AD-A035 883 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 9/2

APL HISTOGRAM, DENSITY ESTIMATION AND PROBABILITY PLOTTING ROUT--ETC(U)

DEC 76 D R HUTCHINSON

UNCLASSIFIED NL

1 OF 2

ADA035883

2

B 5.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

APL HISTOGRAM, DENSITY ESTIMATION
AND
PROBABILITY PLOTTING ROUTINES

by

Dennis Roy Hutchinson

December 1976

Thesis Advisor: P. A. W. Lewis

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) APL Histogram, Density Estimation and Probability Plotting Routines.		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; December 1976
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis Roy Hutchinson		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE December 1976
		13. NUMBER OF PAGES 95
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Data Analysis
(APL) (A Programming Language)
Histograms
Jacknifing
Sectioning
Listing; Sample Statistics
Normal Plotting
Exponential Plotting
Density Estimation
Probability Plotting

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper introduces several data analysis routines that were designed for interactive use with APL (A Programming Language) and placed in the APL user library at the Naval Postgraduate School. Specifically, histograms, density estimation and probability plotting routines are both explained in detail and demonstrated with actual data. In addition, applications and limitations on each of the over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601 |

251 450

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

routines are explored. And, the combined routines give the general user an extensive tool to analyze either discrete or continuous data.

DD Form 1473
1 Jan 73
S/N 0102-014-6601

APL HISTOGRAM, DENSITY ESTIMATION
AND
PROBABILITY PLOTTING ROUTINES

by

Dennis Roy Hutchinson
Captain, United States Army
B.S., United States Military Academy, 1969

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 1976

Author: Dennis R. Hutchinson

Approved by: Peter A. W. Lewis, Thesis Advisor

Richard W. Butterworth, Second Reader

J K Hartman (Acting), Chairman, Department of Operations Research

David A. Schrady, Dean of Information and Policy Sciences

ABSTRACT

This paper introduces several data analysis routines that were designed for interactive use with APL (A Programming Language) and placed in the APL user library at the Naval Postgraduate School. Specifically, histograms, density estimation and probability plotting routines are both explained in detail and demonstrated with actual data. In addition, applications and limitations on each of the routines are explored. And, the combined routines give the general user an extensive tool to analyze either discrete or continuous data.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

ACKNOWLEDGMENTS

Particular acknowledgment is due to two individuals. Professor P.A.W. Lewis, with his extensive knowledge in the area of data analysis, was instrumental in the initial phase of this thesis. And, with his continual guidance the methodology and procedures to be used were successfully implemented in the computer routines created.

Professor Richard W. Butterworth contributed greatly with his extensive knowledge of all aspects of APL. His cooperation and willingness to assist resulted in the efficient and extensive use of all current features available for APL at the Naval Postgraduate School.

I. INTRODUCTION

The Naval Postgraduate School acquired APL (A Programming Language) from IBM in 1974. Since that time more and more students and faculty have become familiar with the extensive and efficient capabilities of APL and have been putting these features to good use. With the acquisition of APL came several extensive library routines that are both well documented and varied in scope. However, on close examination of these library routines it was found that statistics and data analysis were areas where some additions would be particularly useful.

Because of the efficiency and ease of APL in manipulating vectors, matrices and arrays, it is ideal for use in the area of data analysis. After a complete and thorough screening of the existing APL library routines pertaining to data analysis, it was found that by adding six additional data analysis routines to the present library, the Naval Postgraduate School could enhance its present APL capability and provide the student and general user with a more varied and flexible tool for analyzing data.

To this end the purpose of this thesis will be (1) to completely describe the six data analysis routines added to the APL library, (2) to explain the features and capabilities of each of the routines and (3) to demonstrate the use of each of the routines with "real world data".

The data to be used in this paper has come from two different sources. The first source of data was from tests performed jointly by IBM Germany and the German Public Telephone Network on errors in transmission of binary data on telephone lines (Lewis & Cox, 1966). From this source two sets of data are used and each data set contains the times between errors in binary bits transmitted over telephone lines. The first data set contains 672 elements (times-between-errors: actually number of bits between errors) and will hereby be referred to as "telephone data 1". The second data set contains 736 elements and will be referred to as "telephone data 2". The second source of data was obtained from percent overrun or underrun on selected military contracts during the year 1950 (Dixon, 1973). This data set contains 22 elements and will be referred to as "cost overrun data".

II. HISTOGRAM ROUTINE

A. DESCRIPTION

The first routine to be presented is the histogram routine which is used for estimating from given data the probability density function f(x) of a continuous random variable. The current APL library has several small histogram routines that are general in nature but lack the overal detail necessary for good data analysis. For this reason HIST (histogram routine) was created. HIST represents the adaption and modification of the fortran library version of HISTG/F, which was developed at N.P.S. by D. R. Robinson under the guidance of Professor P.A.W. Lewis. By modifying and adapting HISTG/F to APL the power and efficiency of the APL language could be put to full use.

A complete description of how HIST operates is contained in the variable HISTHOW. If the users APL workspace is properly loaded (see section IX.B. for workspace loading procedures) all that is necessary is to type HISTHOW. The user then receives the following printed response on the terminal:

HISTHOW

SYNTAX HIST

HIST ALLOWS YOU TO INTERACTIVELY OBTAIN A HISTOGRAM OF YOUR DATA ALONG WITH A SET OF BASIC DESCRIPTIVE STATISTICS. IN ADDITION, HIST HAS THE FOLLOWING CAPABILITIES WHICH ALLOW YOU:

(1) *THE OPTION OF A TITLE FOR YOUR HISTOGRAM*
(2) *THE OPTION OF DISPLAYING A SMOOTHED EMPIRICAL DENSITY FUNCTION OVER THE HISTOGRAM*
(3) *THE OPTION OF SCALING AND SELECTING THE NUMBER OF CELLS FOR YOUR HISTOGRAM*
(4) *THE OPTION OF SELECTING AN INTERVAL AND PERFORMING A HISTOGRAM ON ALL THE DATA POINTS OR CONDITIONALLY SELECTING AN INTERVAL IN THE RANGE OF THE DATA.*
(5) *THE OPTION OF HAVING YOUR OUTPUT APPEAR ON THE OFFLINE PRINTER OR ON YOUR TERMINAL*

WHEN YOU TYPE HIST YOU WILL BE ASKED TO DO THE FOLLOWING:
(1) *ENTER YOUR DATA IN VECTOR FORM - YOU CAN TYPE YOUR DATA IN SINGLY OR YOU CAN TYPE THE NAME OF A VARIABLE THAT HAS YOUR DATA IN IT. YOU MUST ENSURE THAT YOU HAVE AT LEAST* 10 *DATA POINTS IN YOUR VECTOR AND THAT THERE IS SOME DIFFERENCES IN THE DATA POINTS (MAX SIZE OF INTEGER VECTOR IS APPROX.* 2500 , *MAX SIZE OF REAL VECTOR IS* 2000). *AFTER YOU HAVE ENTERED YOUR DATA YOU WILL BE ASKED*
(2) *IF YOU DESIRE A SMOOTHED EMPIRICAL DENSITY FUNCTION OR NOT. THE EMPIRICAL DENSITY FUNCTION WHEN PLOTTED GIVES ESSENTIALLY A MORE EXACT PICTURE OF THE DATA THAN DOES THE HISTOGRAM ALONE, ALTHOUGH THIS FEATURE IS SLIGHTLY BLURRED BY THE PRECISION WHICH CAN BE OBTAINED WITH THE APL BALL (THE APL FINE PLOT IS NOT PRESENTLY AVAILABLE ON THE NPS SYSTEM). THE SMOOTHED EMPIRICAL DENSITY IS DEFINED BY THE RELATION (LEWIS,LIU,ROBINSON, AND ROSENBLATT,*1975; *ROSENBLATT,*1956)

$$\bar{F}_N(Z) = \frac{1}{N \times B(N)} \sum_{I=1}^{N} W((X_I - Z) \div B(N))$$

WHERE N IS THE NUMBER OF DATA POINTS, B(N) IS A BANDWIDTH FUNCTION,

$$B(N) = RANGE \div SQRT(N)$$

AND W IS A WEIGHT FUNCTION,

$$W(Z) = \begin{cases} 0 & \text{IF } |Z| > 1 \\ 1 - |Z| & \text{OTHERWISE} \end{cases}$$

$\bar{F}(Z)$ IS COMPUTED FOR VALUES OF Z BETWEEN THE MAXIMUM AND THE MINIMUM OF THE SAMPLE AND PLOTTED OVER THE HISTOGRAM USING THE SYMBOL -F-. THE RELATIVE FREQUENCY MARKS ON THE LEFT OF THE OUTPUT REFER TO THE HISTOGRAM, AND NOT TO THE DENSITY FUNCTION. AFTER THIS QUERY YOU WILL BE ASKED

(3) IF YOU DESIRE TO TITLE YOUR HISTOGRAM. IF YOU ELECT TO TITLE YOUR HISTOGRAM, SIMPLY TYPE YOUR TITLE, ENSURING THAT YOUR TITLE IS MORE THAN ONE CHARACTER IN LENGTH. IF NO TITLE IS DESIRED JUST HIT THE CARRIAGE RETURN. AFTER THE TITLE QUERY YOU WILL BE ASKED

(4) IF YOU WANT TO SET YOUR OWN SCALE AND THE NUMBER OF CELLS. YOUR RESPONSE MUST BE A VECTOR OF 3 ELEMENTS THE FIRST ELEMENT IS THE NUMBER OF CELLS YOU DESIRE, THIS MUST BE AN INTEGER BETWEEN 10 AND 28 , THE SECOND ELEMENT IS THE LEFT SCALE POINT AND THE THIRD ELEMENT IS THE RIGHT SCALE POINT (HIST DOES NOT REQUIRE THAT YOUR INTERVAL BE DIVISIBLE BY THE NUMBER OF CELLS). IF YOU WANT HIST TO AUTOMATICALLY SCALE AND PICK THE CELLS YOU SHOULD TYPE THE VECTOR 0 0 0 . AFTER YOU HAVE SELECTED YOUR SCALING TECHNIQUE YOU WILL BE ASKED

(5) IF YOU WANT DATA POINTS NOT INSIDE THE SCALE LIMITS INCLUDED IN THE HISTOGRAM ROUTINE. MOST HISTOGRAMS LUMP DATA POINTS THAT FALL OUTSIDE THE SCALE LIMITS IN THE END CELLS. HOWEVER, HIST GIVES YOU THE OPTION OF INCLUDING THEM OR EXCLUDING THEM, I.E. OF OBTAINING A HISTOGRAM FOR THE CONDITIONAL DENSITY. AFTER YOUR RESPONSE TO THIS QUERY YOU WILL BE ASKED

(6) IF YOU WANT YOUR OUTPUT TO APPEAR ON THE OFFLINE PRINTER OR ON YOUR TERMINAL. IF YOU SELECT THE OFFLINE PRINTER THE NEXT RESPONSE YOU WILL RECEIVE ON YOUR TERMINAL IS - HISTOGRAM SENT TO PRINTER -. THIS RESPONSE WILL TAKE SEVERAL SECONDS AND AFTER IT IS RECEIVED YOUR TERMINAL IS FREE FOR FURTHER USE. HOWEVER, IF YOU ELECTED TO HAVE YOUR HISTOGRAM PRINTED ON YOUR TERMINAL THE PRINTING WOULD BEGIN IN JUST A FEW SECONDS BUT WOULD TAKE BETWEEN 5 AND 10 MINUTES TO COMPLETE.

THE FOLLOWING BASIC DESCRIPTIVE STATISTICS ARE COMPUTED AND PRINTED OUT BY HIST.

MEAN, MEDIAN, TRIMEAN, MIDMEAN, MODE
GEOMETRIC AND HARMONIC MEANS (POSITIVE SAMPLES ONLY)
VARIANCE, STANDARD DEVIATION, COEFFICIENT OF VARIATION, RANGE AND MIDSPREAD
THIRD AND FOURTH CENTRAL MOMENTS, COEFFICIENTS OF SKEWNESS AND KURTOSIS
MAXIMUM, MINIMUM AND 5 SAMPLE QUANTILES

IN ADDITION, THE MEAN IS DISPLAYED ON THE HISTOGRAM BY A VERTICAL COLUMN OF -M- AND THE QUARTILES BY COLUMNS OF DOTS.

INTERPRETING THE OUTPUT

THE DEFINITIONS OF THE BASIC STATISTICS COMPUTED BY HIST ARE LISTED BELOW. PAGE NUMBER REFERENCES ARE TO THE CRC STANDARD MATH TABLES, 19TH EDITION (1971).

MEAN AVERAGE OF THE SAMPLE (P 554).

MEDIAN MID-VALUE OF THE SAMPLE, IF THERE ARE AN ODD NUMBER OF SAMPLE POINTS, OR THE AVERAGE OF THE TWO MIDDLE VALUES FOR AN EVEN NUMBER OF POINTS (P 555)

SAMPLE QUARTILES THE $Q(1)=.25$, $Q(2)=.50$, AND $Q(3)=.75$ POPULATION QUARTILES ARE THE SOLUTION TO THE EQUATION $PROB\ (X \le X(Q(I))) = Q(I)$ $I=1,2,3$. THE SAMPLE QUARTILES, WHICH ESTIMATE THE POPULATION QUARTILES ARE, THE J^{TH} ORDERED VALUE IN THE SAMPLE, WHERE $J = [\ Q(I) \times N\ ] + 1$. WHERE N = SAMPLE SIZE.

TRIMEAN $0.25 \times (Q(1) + 2Q(2) + Q(3))$, WHERE THE $Q(I)$ ARE THE QUARTILES.

MIDMEAN THE AVERAGE OF ALL THE SAMPLE VALUES BETWEEN THE UPPER AND LOWER QUARTILES.

MODE THE DATA POINT THAT OCCURS MOST OFTEN (IF ALL THE DATA POINTS ARE DIFFERENT OR IF THERE ARE MORE THAN 300 DATA POINTS THE MODE WILL NOT BE PRINTED. IF TWO OR MORE MODES OCCUR HIST WILL PRINT THE FIRST MODE.)

MIDRANGE AVERAGE OF THE MAXIMUM AND MINIMUM.

GEOMETRIC MEAN (P 554).

HARMONIC MEAN (P 555).

VARIANCE (P 557). UNBIASED ESTIMATORS FOR VARIANCE AND STANDARD DEVIATION ARE USED.

STANDARD DEVIATION (P 557).

COEFFICIENT OF VARIATION = STANDARD DEVIATION ÷ |MEAN| WHEN THE MEAN IS LESS THAN 1E-30, THE COEFFICIENT OF VARIATION IS SET TO ZERO.

MEAN DEVIATION (P 556). *THE AVERAGE OF THE SUM OF THE ABSOLUTE DIFFERENCES BETWEEN THE SAMPLE VALUES AND THE MEDIAN.*

RANGE *MAXIMUM - MINIMUM* (P 557).

MIDSPREAD Q(3) - Q(1) , *ALSO CALLED THE INTERQUARTILE DISTANCE.*

M3 *THIRD CENTRAL MOMENT. UNBIASED ESTIMATOR IS USED.* (P 558)

M4 *FOURTH CENTRAL MOMENT. UNBIASED ESTIMATOR IS USED.* (P 558)

COEFFICIENT OF SKEWNESS *M3 ÷ (STD DEV)*3*

COEFFICIENT OF KURTOSIS (*M4 ÷ (STD DEV)*4*) - 3

BETA1 *BIASED ESTIMATE OF THIRD CENTRAL MOMENT. CAN BE USED IN TESTING FOR NORMALITY. (BIOMETRIKA TABLES FOR STATISTICIANS,*1966).

BETA2 *BIASED ESTIMATE OF FOURTH CENTRAL MOMENT. (BIOMETRIKA TABLES FOR STATISTICIANS,*1966).

MAXIMUM *LARGEST SAMPLE VALUE.*

MINIMUM *SMALLEST SAMPLE VALUE.*

SAMPLE QUANTILES *THE* α*-QUANTILE,* $X(\alpha)$*, IS THE SOLUTION TO THE EQ. PROBABILITY* $(X \leq X(\alpha)) = \alpha$.

With this complete description the general user should be able to take full advantage of HIST and put to use all its options.

B. USAGE WITH TELEPHONE DATA 1 AND TELEPHONE DATA 2, OFFLINE, ALL DATA, ECDF, AND TITLE

HIST was now used on two sets of data. Both telephone data 1 and telephone data 2 were first used with the offline printer demonstrating the title option, the empirical density function option and using the conditional option with any data points outside the designated interval being lumped into the end cells. When HIST was typed the following responses to each of the queries were entered.

```
      HIST
ENTER DATA IN VECTOR FORM
⎕:
      TELDAT1

IF YOU ALSO WANT A SMOOTHED EMPIRICAL DENSITY FUNCTION ENTER
A    1   .  IF YOU DO NOT WANT IT ENTER A    0   .
⎕:
      1

IF  YOU WANT  TO  TITLE  YOUR  HISTOGRAM  TYPE  YOUR  TITLE.
IF  YOU  DO  NOT WANT A TITLE JUST HIT THE CARRIAGE  RETURN.

TELEPHONE DATA 1

IF YOU WANT TO SET THE NUMBER OF CELLS AND THE  SCALE  ENTER
FIRST THE  NUMBER OF CELLS (AN INTEGER BETWEEN  10  AND  28)
FOLLOWED BY A SPACE AND THEN  YOUR LEFT SCALE POINT FOLLOWED
BY A SPACE AND THEN YOUR RIGHT SCALE POINT.  HOWEVER, IF YOU
WANT HIST TO AUTOMATICALLY SCALE ENTER   0 0 0   .
⎕:
      28 0 20000

GIVEN THAT YOU HAVE SET YOUR  OWN  SCALE,  TO  INCLUDE  DATA
POINTS THAT MIGHT  BE OUTSIDE YOUR SCALE LIMITS  IN  THE END
CELLS, TYPE   1   .  IF YOU DESIGNATED  AUTOSCALE  ALSO, TYPE
1   .  IF HOWEVER, YOU DO NOT WANT THE DATA OUTSIDE THE SCALE
LIMITS INCLUDED IN THE HISTOGRAM, TYPE   0   .
⎕
      1

IF YOU WANT YOUR OUTPUT TO APPEAR ON  THE  OFFLINE  PRINTER,
TYPE   1   .  IF YOU  WANT YOUR  OUTPUT  TO  APPEAR  ON  YOUR
TERMINAL, TYPE   0   .  (NOTE IF YOU TYPED   0  BE SURE YOUR
TERMINALS CARRIAGE PAGE SETTING IS  ON  THE  MAXIMUM  WIDTH)
⎕:
      1
HISTOGRAM SENT TO PRINTER
```

Note that telephone data 1 was contained in the variable TELDAT1 and that the number of cells chosen was 28 with the left scale point being 0 and the right scale point being 20,000.

After the response - HISTOGRAM SENT TO PRINTER - was received. HIST was again typed under identical conditions and telephone data 2 was entered through the variable TELDAT2.

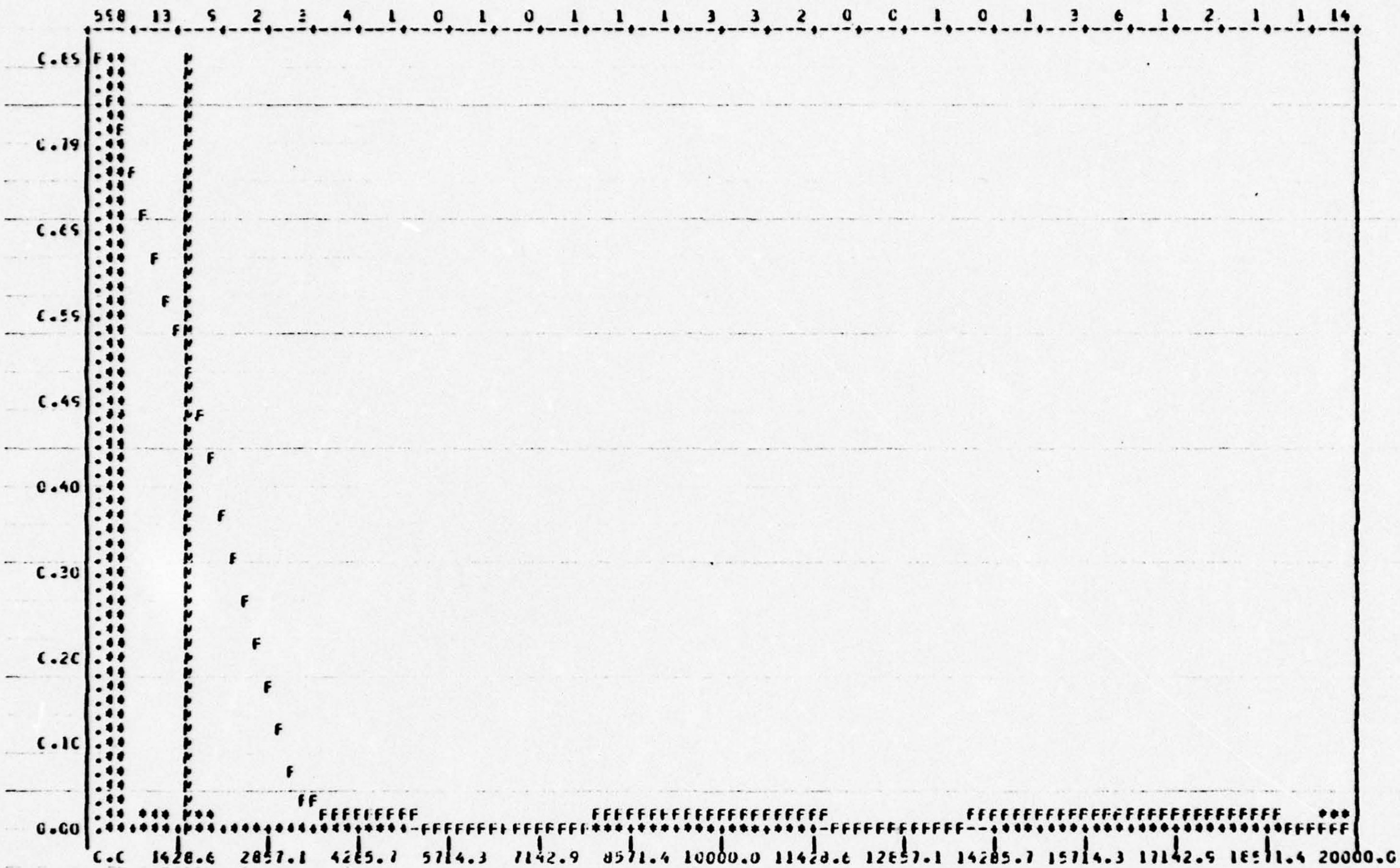
```
      HIST
ENTER DATA IN VECTOR FORM
⎕:
      TELDAT2

IF YOU ALSO WANT A SMOOTHED EMPIRICAL DENSITY FUNCTION ENTER
A   1  .  IF YOU DO NOT WANT IT ENTER A   0  .
⎕:
      1

IF  YOU WANT  TO  TITLE  YOUR  HISTOGRAM  TYPE  YOUR  TITLE.
IF  YOU  DO  NOT WANT A TITLE JUST HIT THE CARRIAGE  RETURN.

TELEPHONE DATA 2

IF YOU WANT TO SET THE NUMBER OF CELLS AND THE  SCALE  ENTER
FIRST THE  NUMBER OF CELLS (AN INTEGER BETWEEN  10  AND  28)
FOLLOWED BY A SPACE AND THEN  YOUR LEFT SCALE POINT FOLLOWED
BY A SPACE AND THEN YOUR RIGHT SCALE POINT.  HOWEVER, IF YOU
WANT HIST TO AUTOMATICALLY SCALE ENTER   0 0 0  .
⎕:
      28 0 20000

GIVEN THAT YOU HAVE SET YOUR  OWN  SCALE,  TO  INCLUDE  DATA
POINTS THAT MIGHT  BE OUTSIDE YOUR SCALE LIMITS  IN  THE END
CELLS, TYPE   1  .  IF YOU DESIGNATED  AUTOSCALE  ALSO, TYPE
1  .  IF HOWEVER, YOU DO NOT WANT THE DATA OUTSIDE THE SCALE
LIMITS INCLUDED IN THE HISTOGRAM, TYPE   0  .
⎕:
      1

IF YOU WANT YOUR OUTPUT TO APPEAR ON  THE  OFFLINE  PRINTER,
TYPE   1  .  IF YOU  WANT YOUR  OUTPUT  TO  APPEAR  ON  YOUR
TERMINAL, TYPE   0  .  (NOTE IF YOU TYPED   0  BE SURE YOUR
TERMINALS CARRIAGE PAGE SETTING IS  ON  THE  MAXIMUM  WIDTH)
⎕:
      1
HISTOGRAM SENT TO PRINTER
```

Now by looking at figure 1 (output for telephone data 1) and figure 2 (output from telephone data 2) the similarities and differences in the histograms can be compared. Without getting into specifics, the empirical density function plot seems to indicate that both sets of data are similar. However, the one time-between-errors dominate the data; a more detailed discussion of this data and its analysis is given in Section VIII.

FIGURE 1

TELEPHONE DATA 1

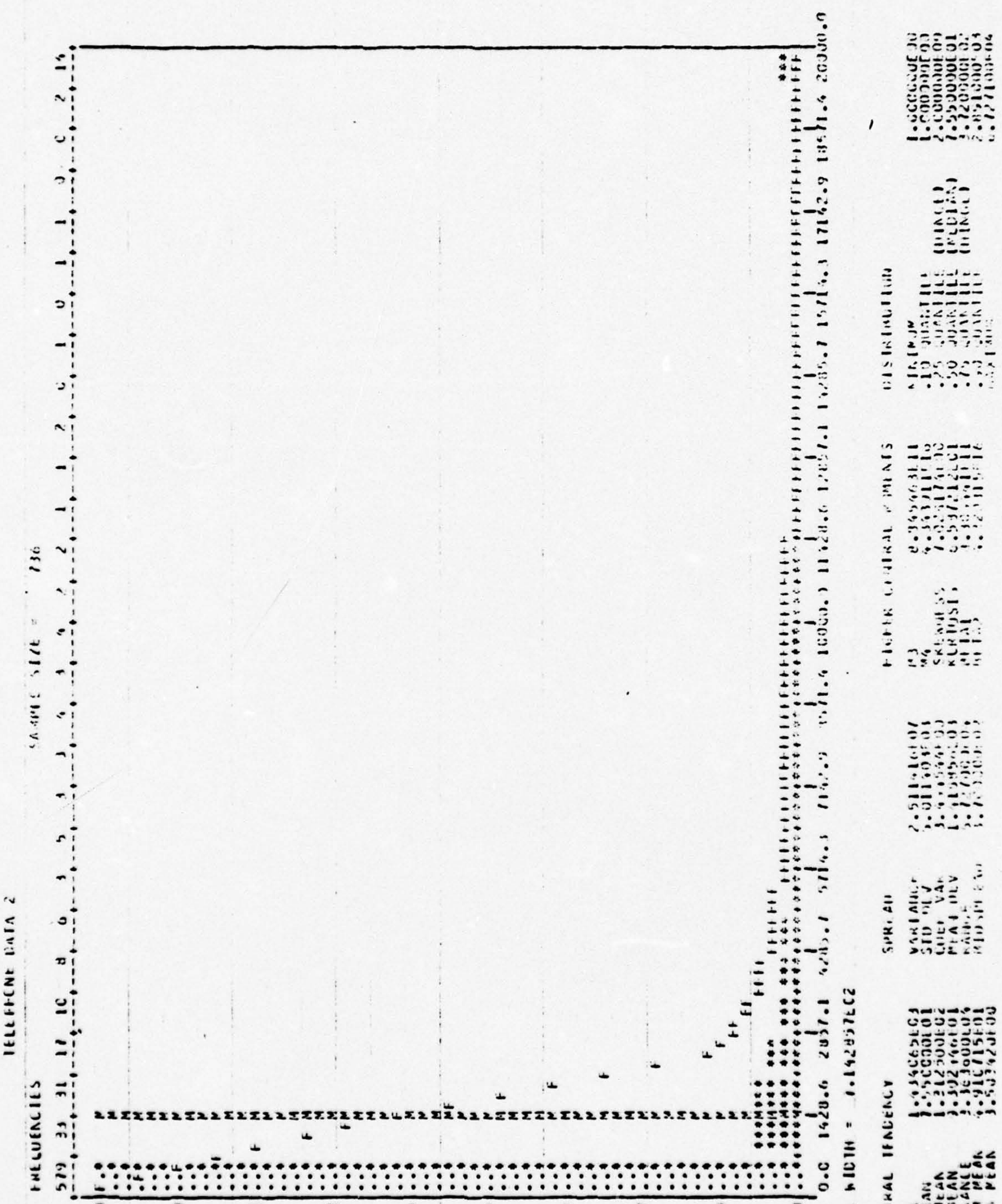
FREQUENCIES   SAMPLE SIZE = 672



CELL WIDTH = 7.142857E02

CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION	
MEAN	1.546215E03	VARIANCE	4.836184E07	M3	2.405748E12	MINIMUM	1.000000E00
MEDIAN	1.400000E01	STD DEV	6.554268E03	M4	1.534495E17	.10 QUANTILE	1.000000E00
TRIMEAN	3.050000E01	COEF VAR	4.491786E00	SKEWNESS	7.153127E00	.25 QUANTILE (HINGE)	2.000000E00
MIDMEAN	2.215540E01	MEAN DEV	1.544295E03	KURTOSIS	6.260848E01	.50 QUANTILE (MEDIAN)	1.400000E01
MIDRANGE	4.295700E04	RANGE	8.559200E04	BETA1	2.355015E12	.75 QUANTILE (HINGE)	5.200000E01
GEOM MEAN	2.137672E01	MIDSPREAD	5.000000E01	BETA2	1.525570E17	.90 QUANTILE	1.270000E03
HARM MEAN	3.541554E00					MAXIMUM	8.559300E04

FIGURE 2



C. USAGE WITH TELEPHONE DATA 1 AND TELEPHONE DATA 2, ON LINE; CONDITIONAL DATA BETWEEN 2 AND 140, ECDF, AND TITLE

Because both sets of data contain:

(1) a large number of elements,

(2) a large number of times-between-error equal to 1 (this becomes more apparent when HISTLIST is described), and

(3) the range of the data sets is so extensive,

it would appear that the conditional option available on HIST could be used to see if the two data sets are in fact similar over a smaller interval. This in fact was done using the on line printer option, the empirical density function option, the title option and the conditional option with any data points outside the designated interval excluded from the histogram.

```
      HIST
ENTER DATA IN VECTOR FORM
□:
      TELDAT1

IF YOU ALSO WANT A SMOOTHED EMPIRICAL DENSITY FUNCTION ENTER
A   1  . IF YOU DO NOT WANT IT ENTER A   0  .
□:
      1

IF  YOU WANT  TO  TITLE  YOUR  HISTOGRAM  TYPE  YOUR  TITLE.
IF  YOU  DO  NOT WANT A TITLE JUST HIT THE CARRIAGE  RETURN.

TELEPHONE DATA 1 BETWEEN 2 AND 140

IF YOU WANT TO SET THE NUMBER OF CELLS AND THE  SCALE  ENTER
FIRST THE  NUMBER OF CELLS (AN INTEGER BETWEEN  10  AND  28)
FOLLOWED BY A SPACE AND THEN  YOUR LEFT SCALE POINT FOLLOWED
BY A SPACE AND THEN YOUR RIGHT SCALE POINT.  HOWEVER, IF YOU
WANT HIST TO AUTOMATICALLY SCALE ENTER   0 0 0  .
□:
      28 2 140
```

```
GIVEN THAT YOU HAVE SET YOUR OWN SCALE, TO INCLUDE DATA
POINTS THAT MIGHT BE OUTSIDE YOUR SCALE LIMITS IN THE END
CELLS, TYPE  1  . IF YOU DESIGNATED AUTOSCALE ALSO, TYPE
1  . IF HOWEVER, YOU DO NOT WANT THE DATA OUTSIDE THE SCALE
LIMITS INCLUDED IN THE HISTOGRAM, TYPE  0  .
⎕:
      0

IF YOU WANT YOUR OUTPUT TO APPEAR ON THE OFFLINE PRINTER,
TYPE  1  . IF YOU WANT YOUR OUTPUT TO APPEAR ON YOUR
TERMINAL, TYPE  0  . (NOTE IF YOU TYPED  0  BE SURE YOUR
TERMINALS CARRIAGE PAGE SETTING IS ON THE MAXIMUM WIDTH)
⎕:
      0
```

Note that the same variable TELDAT1 is used but this time the interval was between 2 and 140. Also, the - HISTOGRAM SENT TO PRINTER - was not typed because the on-line printer (terminal) option was employed.

After the output for telephone data 1 was printed HIST was again typed and telephone data 2 was entered under identical conditions.

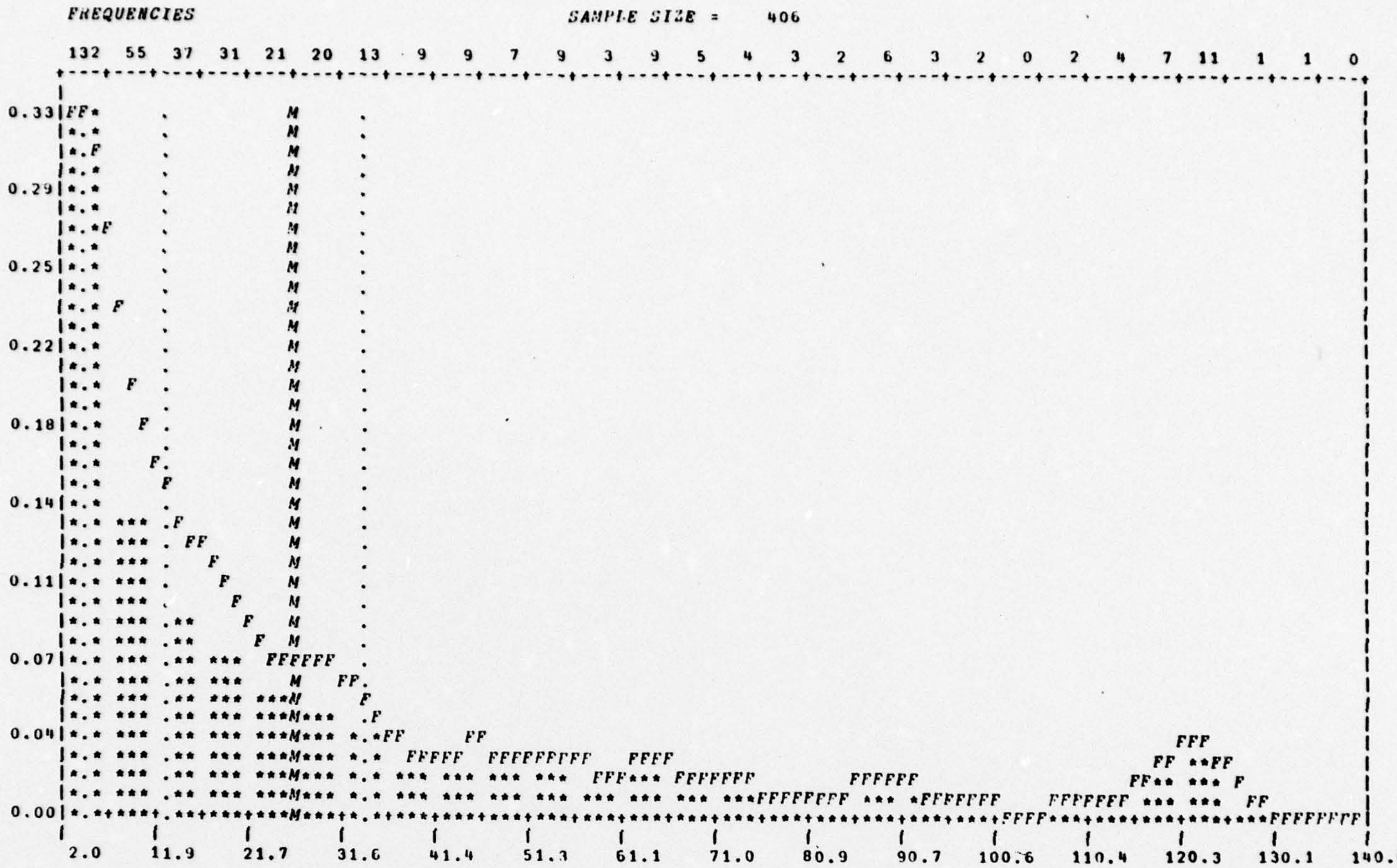
```
      HIST
ENTER DATA IN VECTOR FORM
⎕:
      TELDAT2

IF YOU ALSO WANT A SMOOTHED EMPIRICAL DENSITY FUNCTION ENTER
A  1  . IF YOU DO NOT WANT IT ENTER A  0  .
⎕:
      1

IF YOU WANT TO TITLE YOUR HISTOGRAM TYPE YOUR TITLE.
IF YOU DO NOT WANT A TITLE JUST HIT THE CARRIAGE RETURN.

TELEPHONE DATA 2 BETWEEN 2 AND 140

IF YOU WANT TO SET THE NUMBER OF CELLS AND THE SCALE ENTER
FIRST THE NUMBER OF CELLS (AN INTEGER BETWEEN  10  AND  28)
FOLLOWED BY A SPACE AND THEN YOUR LEFT SCALE POINT FOLLOWED
BY A SPACE AND THEN YOUR RIGHT SCALE POINT. HOWEVER, IF YOU
WANT HIST TO AUTOMATICALLY SCALE ENTER  0 0 0  .
⎕:
      28 2 140
```

```
GIVEN THAT YOU HAVE SET YOUR OWN SCALE, TO INCLUDE DATA
POINTS THAT MIGHT BE OUTSIDE YOUR SCALE LIMITS IN THE END
CELLS, TYPE  1  . IF YOU DESIGNATED AUTOSCALE ALSO, TYPE
1  . IF HOWEVER, YOU DO NOT WANT THE DATA OUTSIDE THE SCALE
LIMITS INCLUDED IN THE HISTOGRAM, TYPE  0  .
⎕:
      0

IF YOU WANT YOUR OUTPUT TO APPEAR ON THE OFFLINE PRINTER,
TYPE  1  . IF YOU WANT YOUR OUTPUT TO APPEAR ON YOUR
TERMINAL, TYPE  0  . (NOTE IF YOU TYPED  0  BE SURE YOUR
TERMINALS CARRIAGE PAGE SETTING IS ON THE MAXIMUM WIDTH)
⎕:
      0
```

Figure 3 (output from telephone data 1 between 2 and 140) and figure 4 (output from telephone data 2 between 2 and 140) now appear quite different in shape based on the empirical density function plot. This is, again, because of the extensive range of the data (85,993 for telephone data 1 and 67,271 for telephone data 2) and the large number of times-between-error equal to one. Both sets of data are actually discrete, only occurring at multiples of 1, but as an initial analysis the data sets were treated as continuous. Thus, by employing the conditional option available on HIST differences in the two sets of data become quite apparent whereas before, the differences were not so easily detected.

FIGURE 3

TELEPHONE DATA 1 BETWEEN 2 AND 140

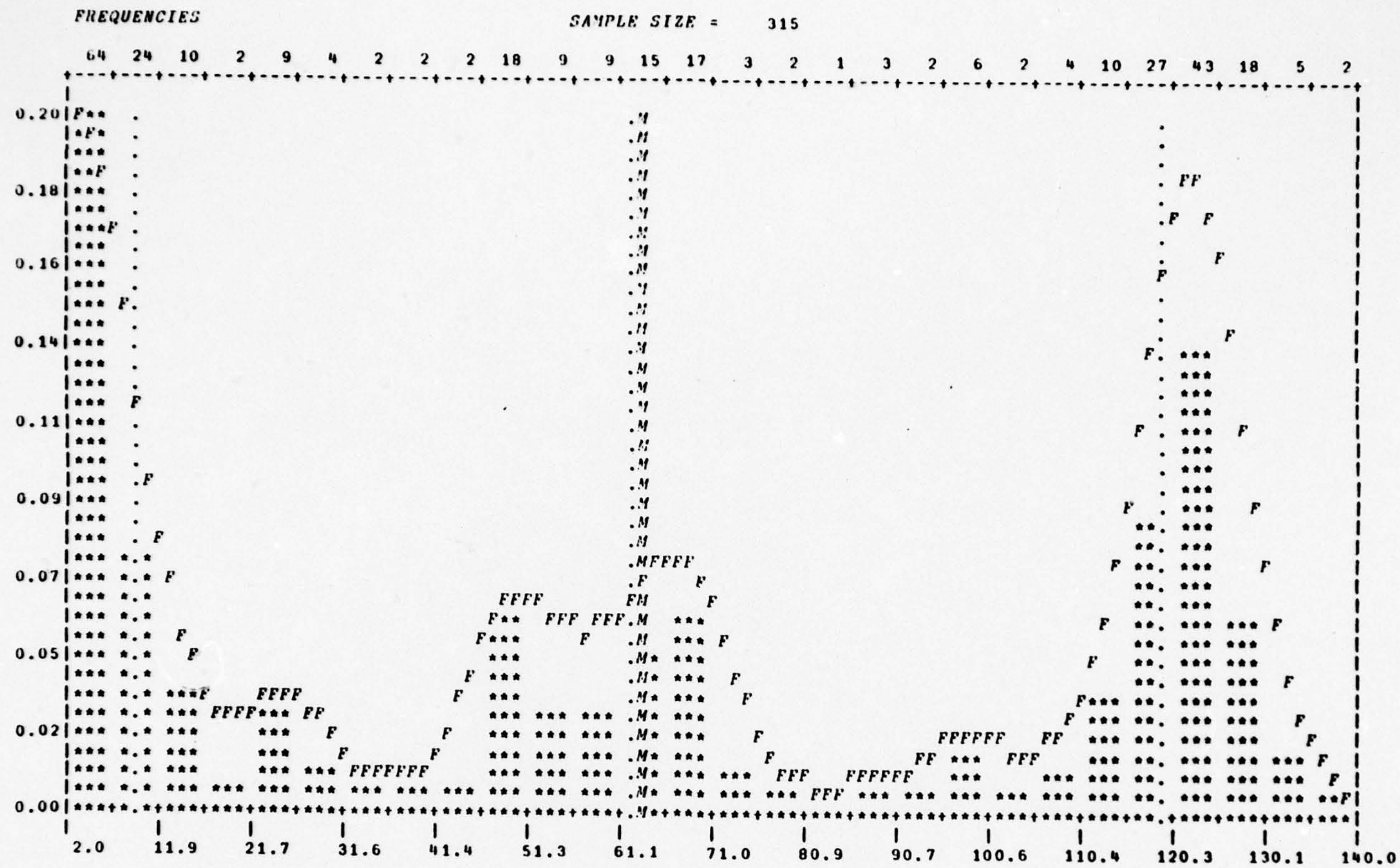


CELL WIDTH = 4.928571E00

CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION	
MEAN	2.716995E01	VARIANCE	1.070057E03	M3	5.920433E04	MINIMUM	2.000000E00
MEDIAN	1.300000E01	STD DEV	3.271173E01	M4	5.680039E06	.10 QUANTILE	2.000000E00
TRIMEAN	1.600000E01	COEF VAR	1.203967E00	SKEWNESS	1.691386E00	.25 QUANTILE (HINGE)	4.000000E00
MIDMEAN	1.533824E01	MEAN DEV	2.165271E01	KURTOSIS	1.960633E00	.50 QUANTILE (MEDIAN)	1.300000E01
MIDRANGE	6.850000E01	RANGE	1.330000E02	BETA1	5.876758E04	.75 QUANTILE (HINGE)	3.400000E01
GEOM MEAN	1.319364E01	MIDSPREAD	3.000000E01	BETA2	5.640937E06	.90 QUANTILE	7.900000E01
HARM MEAN	6.639325E00					MAXIMUM	1.350000E02

FIGURE 4

TELEPHONE DATA 2 BETWEEN 2 AND 140



CELL WIDTH = 4.928571E00

CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION	
MEAN	6.410159E01	VARIANCE	2.397850E03	M3	3.892334E03	MINIMUM	2.000000E00
MEDIAN	6.200000E01	STD DEV	4.896784E01	M4	8.039527E06	.10 QUANTILE	2.000000E00
TRIMEAN	6.325000E01	COEF VAR	7.639100E-01	SKEWNESS	3.314952E-02	.25 QUANTILE (HINGE)	1.000000E01
MIDMEAN	6.393082E01	MEAN DEV	4.372698E01	KURTOSIS	-1.601744E00	.50 QUANTILE (MEDIAN)	6.200000E01
MIDRANGE	7.100000E01	RANGE	1.380000E02	BETA1	3.855343E03	.75 QUANTILE (HINGE)	1.190000E02
GEOM MEAN	3.273448E01	MIDSPREAD	1.090000E02	BETA2	8.046434E06	.90 QUANTILE	1.250000E02
HARM MEAN	9.938929E00					MAXIMUM	1.400000E02

III. LISTING ROUTINE

A. DESCRIPTION

The second routine presented is a listing routine. APL has a function that will automatically sort the data and print the results. However, the unique feature of HISTLIST (listing routine) is that it takes advantage of like occurrences in the data and prints the ordered data ascendingly in a compressed form. This becomes highly useful when listing a large number of data points that contain multiple occurrences. It is also a tool for finding multiplicities in supposedly continuous data, and a probability function estimating routine for data which is known to be discrete.

A complete description of how HISTLIST operates is contained in the variable HISTLISTHOW. When the user types HISTLISTHOW the following response is printed on the terminal:

```
      HISTLISTHOW

SYNTAX  HISTLIST

    HISTLIST  IS A HIGHLY CONVENIENT WAY TO LIST  YOUR  DATA.
HISTLIST TAKES YOUR DATA, ORDERS IT AND COMPRESSES  IT.  FOR
EXAMPLE, IF THREE  DATA  POINTS  WERE  ALL  THE  SAME  VALUE
HISTLIST WOULD JUST PRINT THE VALUE ONCE AND THEN PRINT  THE
NUMBER OF OCCURENCES  OF  THAT  VALUE.  HISTLIST  WILL  ALSO
PRINT THE SERIAL NUMBER OF  THE  DATA, THE  PERCENTAGE  THIS
SAMPLE VALUE IS TO THE WHOLE SAMPLE, AND A  SMALL  HISTOGRAM
(STARS) SHOWING RELATIVE PERCENTAGES. EXAMPLE: 6  4  4  3  4

HISTLIST
```

SER. NUM.	ORDERED DATA	NUMBER OF OCCURENCES		PER CENT
1	3	1	****	.20
2	4	3	************	.60
5	6	1	****	.20

HISTLIST IS IDEALLY SUITED FOR A LARGE SAMPLE THAT COULD POSSIBLY HAVE A LOT OF LIKE OCCURENCES. HISTLIST FURTHER HAS THE ADVANTAGE OF BEING USED WITH EITHER THE OFFLINE PRINTER OR THE USERS TERMINAL.

B. USAGE WITH TELEPHONE DATA 1 AND TELEPHONE DATA 2 OFFLINE

HISTLIST was used with the title option and offline printer option on both telephone data 1 and telephone data 2. When HISTLIST was typed the following responses to each of the queries were entered.

```
        HISTLIST
HISTLIST PRINTS THE SERIAL NUMBER OF THE COMPRESSED
DATA, THE ORDERED DATA COMPRESSED, AND THE NUMBER OF
LIKE OCCURENCES. ENTER YOUR DATA IN VECTOR FORM.
⎕:
        TELDAT1

IF YOU WANT TO TITLE YOUR DATA TYPE YOUR TITLE.
IF YOU DO NOT WANT A TITLE JUST HIT THE CARRIAGE
RETURN.

TELEPHONE DATA 1

IF YOU WANT YOUR OUTPUT TO APPEAR ON THE OFFLINE
PRINTER TYPE 1 . IF YOU WANT YOUR OUTPUT TO APPEAR
ON YOUR TERMINAL TYPE 0 .
⎕:
        1
HISTLIST SENT TO PRINTER
```

After the response - HISTLIST SENT TO PRINTER - was received HISTLIST was again typed and telephone data 2 was entered.

```
HISTLIST
HISTLIST PRINTS THE SERIAL NUMBER OF THE  COMPRESSED
DATA, THE ORDERED DATA COMPRESSED, AND THE NUMBER OF
LIKE  OCCURENCES.  ENTER YOUR  DATA  IN VECTOR FORM.
⎕:
      TELDAT2

IF YOU WANT TO  TITLE  YOUR DATA  TYPE  YOUR  TITLE.
IF YOU DO NOT WANT A TITLE  JUST  HIT  THE  CARRIAGE
RETURN.

TELEPHONE DATA 2

IF YOU WANT YOUR  OUTPUT TO APPEAR  ON  THE  OFFLINE
PRINTER TYPE  1   . IF YOU WANT YOUR OUTPUT TO APPEAR
ON YOUR TERMINAL TYPE  0   .
⎕:
      1
HISTLIST SENT TO PRINTER
```

Looking at figure 5 (output with telephone data 1) and figure 6 (output with telephone data 2) the listings of the two data sets can be compared. It can be seen that both telephone data 1 and telephone data 2 contain a large number of multiple occurrences of the number one and the number two. In fact 19% of telephone data 1 is the number one and 24% of telephone data 2 is the number one. Also, telephone data 2 has many more multiple occurrences in the 120 to 130 range than telephone data 1. This was quickly apparent when one looked at the stars to the right of the ordered data.

FIGURE 5A

~~TELEPHONE DATA 1~~

SERIAL NUMBER	ORDERED DATA	NUMBER OF OCCURENCES		PER CENT
1	1.000000	128	************	0.190
129	2.000000	54	*****	0.080
183	3.000000	28	***	0.042
211	4.000000	22	**	0.033
233	5.000000	17	**	0.025
250	6.000000	11	*	0.016
261	7.000000	10	*	0.015
271	8.000000	12	*	0.018
283	9.000000	14	*	0.021
297	10.000000	9	*	0.013
306	11.000000	10	*	0.015
316	12.000000	11	*	0.016
327	13.000000	6	*	0.009
333	14.000000	6	*	0.009
339	15.000000	6	*	0.009
345	16.000000	8	*	0.012
353	17.000000	8	*	0.012
361	18.000000	5		0.007
366	19.000000	12	*	0.018
378	20.000000	1		0.001
379	21.000000	5		0.007
384	22.000000	5		0.007
389	23.000000	3		0.004
392	24.000000	7	*	0.010
399	25.000000	3		0.004
402	26.000000	3		0.004
405	27.000000	2		0.003
407	28.000000	3		0.004
410	29.000000	5		0.007
415	30.000000	6	*	0.009
421	31.000000	4		0.006
425	32.000000	4		0.006
429	33.000000	2		0.003
431	34.000000	4		0.006
435	35.000000	3		0.004
438	37.000000	2		0.003
440	38.000000	2		0.003
442	39.000000	1		0.001
443	40.000000	2		0.003
445	41.000000	2		0.003
447	43.000000	1		0.001
448	44.000000	1		0.001
449	45.000000	4		0.006
453	46.000000	3		0.004
456	47.000000	1		0.001
457	48.000000	2		0.003
459	49.000000	1		0.001
460	50.000000	1		0.001
461	51.000000	2		0.003
463	52.000000	1		0.001
464	53.000000	1		0.001
465	54.000000	3		0.004
468	55.000000	2		0.003
470	56.000000	2		0.003
472	58.000000	2		0.003
474	59.000000	1		0.001
475	62.000000	2		0.003
477	63.000000	3		0.004
480	64.000000	1		0.001
481	65.000000	2		0.003
483	66.000000	1		0.001
484	68.000000	3		0.004
487	69.000000	1		0.001
488	70.000000	1		0.001
489	73.000000	1		0.001
490	74.000000	2		0.003
492	75.000000	1		0.001
493	79.000000	3		0.004
496	83.000000	1		0.001
497	84.000000	1		0.001
498	86.000000	1		0.001
499	88.000000	2		0.003
501	89.000000	1		0.001
502	90.000000	2		0.003
504	91.000000	1		0.001
505	93.000000	1		0.001
506	95.000000	1		0.001
507	98.000000	1		0.001
508	99.000000	1		0.001
509	106.000000	1		0.001
510	109.000000	1		0.001
511	111.000000	1		0.001
512	112.000000	2		0.003
514	113.000000	1		0.001
515	116.000000	1		0.001
516	117.000000	1		0.001
517	119.000000	1		0.001

FIGURE 5B

518	120.000000	4	0.006
522	121.000000	2	0.003
524	122.000000	2	0.003
526	123.000000	4	0.006
530	124.000000	3	0.004
533	128.000000	1	0.001
534	135.000000	1	0.001
535	142.000000	1	0.001
536	148.000000	1	0.001
537	153.000000	1	0.001
538	156.000000	1	0.001
539	158.000000	1	0.001
540	161.000000	1	0.001
541	165.000000	1	0.001
542	175.000000	1	0.001
543	176.000000	1	0.001
544	177.000000	1	0.001
545	183.000000	1	0.001
546	186.000000	1	0.001
547	187.000000	1	0.001
548	192.000000	1	0.001
549	193.000000	1	0.001
550	202.000000	1	0.001
551	217.000000	2	0.003
553	224.000000	1	0.001
554	226.000000	1	0.001
555	228.000000	3	0.004
558	231.000000	1	0.001
559	234.000000	1	0.001
560	237.000000	1	0.001
561	239.000000	2	0.003
563	240.000000	4	0.006
567	241.000000	1	0.001
568	244.000000	1	0.001
569	248.000000	1	0.001
570	249.000000	1	0.001
571	251.000000	1	0.001
572	252.000000	1	0.001
573	270.000000	1	0.001
574	279.000000	1	0.001
575	286.000000	1	0.001
576	297.000000	1	0.001
577	303.000000	1	0.001
578	312.000000	1	0.001
579	318.000000	1	0.001
580	347.000000	1	0.001
581	354.000000	1	0.001
582	360.000000	1	0.001
583	364.000000	1	0.001
584	370.000000	1	0.001
585	390.000000	1	0.001
586	394.000000	1	0.001
587	435.000000	1	0.001
588	460.000000	1	0.001
589	465.000000	1	0.001
590	473.000000	1	0.001
591	480.000000	1	0.001
592	491.000000	1	0.001
593	549.000000	1	0.001
594	600.000000	2	0.003
596	609.000000	1	0.001
597	621.000000	1	0.001
598	711.000000	1	0.001
599	817.000000	1	0.001
600	836.000000	1	0.001
601	927.000000	1	0.001
602	1124.000000	1	0.001
603	1150.000000	1	0.001
604	1270.000000	1	0.001
605	1289.000000	1	0.001
606	1298.000000	1	0.001
607	1305.000000	1	0.001
608	1328.000000	1	0.001
609	1348.000000	1	0.001
610	1355.000000	1	0.001
611	1412.000000	1	0.001
612	1429.000000	1	0.001
613	1489.000000	1	0.001
614	1493.000000	1	0.001
615	1510.000000	1	0.001
616	1519.000000	1	0.001
617	1547.000000	1	0.001
618	1633.000000	1	0.001
619	1787.000000	1	0.001
620	2072.000000	1	0.001
621	2483.000000	1	0.001
622	2806.000000	1	0.001
623	2962.000000	1	0.001
624	3026.000000	1	0.001
625	3201.000000	1	0.001
626	3593.000000	1	0.001
627	3685.000000	1	0.001
628	3952.000000	1	0.001
629	4157.000000	1	0.001
630	4469.000000	1	0.001

FIGURE 5C

631	6208.000000	1	0.001
632	7614.000000	1	0.001
633	8322.000000	1	0.001
634	9015.000000	1	0.001
635	9625.000000	1	0.001
636	9606.000000	1	0.001
637	9818.000000	1	0.001
638	10154.000000	1	0.001
639	10398.000000	1	0.001
640	10451.000000	1	0.001
641	10939.000000	1	0.001
642	11280.000000	1	0.001
643	13447.000000	1	0.001
644	14365.000000	1	0.001
645	15155.000000	1	0.001
646	15294.000000	1	0.001
647	15504.000000	1	0.001
648	15847.000000	1	0.001
649	15868.000000	1	0.001
650	16280.000000	1	0.001
651	16299.000000	1	0.001
652	16361.000000	1	0.001
653	16408.000000	1	0.001
654	16817.000000	1	0.001
655	17174.000000	1	0.001
656	17667.000000	1	0.001
657	18218.000000	1	0.001
658	18649.000000	1	0.001
659	19461.000000	1	0.001
660	21048.000000	1	0.001
661	23498.000000	1	0.001
662	24692.000000	1	0.001
663	26443.000000	1	0.001
664	30974.000000	1	0.001
665	35644.000000	1	0.001
666	38003.000000	1	0.001
667	40131.000000	1	0.001
668	47120.000000	1	0.001
669	47592.000000	1	0.001
670	61710.000000	1	0.001
671	69175.000000	1	0.001
672	85993.000000	1	0.001

FIGURE 6A

TELEPHONE DATA 2

SERIAL NUMBER	ORDERED DATA	NUMBER OF OCCURENCES		PER CENT
1	1.000000	178	****************	0.242
179	2.000000	36	***	0.049
215	3.000000	11	*	0.015
226	4.000000	6		0.008
232	5.000000	6		0.008
238	6.000000	5		0.007
243	7.000000	5		0.007
248	8.000000	4		0.005
252	9.000000	4		0.005
256	10.000000	9	*	0.012
265	11.000000	2		0.003
267	12.000000	3		0.004
270	13.000000	1		0.001
271	14.000000	1		0.001
272	15.000000	1		0.001
273	16.000000	4		0.005
277	19.000000	1		0.001
278	21.000000	1		0.001
279	22.000000	1		0.001
280	24.000000	3		0.004
283	25.000000	3		0.004
286	26.000000	2		0.003
288	27.000000	1		0.001
289	30.000000	3		0.004
292	32.000000	1		0.001
293	36.000000	1		0.001
294	40.000000	2		0.003
296	42.000000	1		0.001
297	44.000000	1		0.001
298	47.000000	4		0.005
302	48.000000	1		0.001
303	49.000000	9	*	0.012
312	50.000000	4		0.005
316	53.000000	3		0.004
319	55.000000	2		0.003
321	56.000000	4		0.005
325	57.000000	3		0.004
328	58.000000	1		0.001
329	59.000000	3		0.004
332	60.000000	2		0.003
334	62.000000	3		0.004
337	63.000000	1		0.001
338	64.000000	4		0.005
342	65.000000	3		0.004
345	66.000000	4		0.005
349	67.000000	1		0.001
350	68.000000	5		0.007
355	69.000000	3		0.004
358	70.000000	3		0.004
361	71.000000	5		0.007
366	73.000000	1		0.001
367	74.000000	1		0.001
368	75.000000	1		0.001
369	76.000000	1		0.001
370	77.000000	1		0.001
371	81.000000	1		0.001
372	86.000000	1		0.001
373	88.000000	1		0.001
374	89.000000	1		0.001
375	93.000000	1		0.001
376	95.000000	1		0.001
377	96.000000	1		0.001
378	97.000000	1		0.001
379	99.000000	3		0.004
382	100.000000	1		0.001
383	101.000000	1		0.001
384	105.000000	1		0.001
385	106.000000	1		0.001
386	107.000000	1		0.001
387	108.000000	1		0.001
388	110.000000	1		0.001
389	111.000000	1		0.001
390	112.000000	1		0.001
391	113.000000	4		0.005
395	114.000000	3		0.004
398	115.000000	1		0.001
399	116.000000	4		0.005
403	117.000000	1		0.001
404	118.000000	5		0.007
409	119.000000	8	*	0.011
417	120.000000	9	*	0.012
426	121.000000	7	*	0.010
433	122.000000	12	*	0.016
445	123.000000	8	*	0.011
453	124.000000	7	*	0.010
460	125.000000	9	*	0.012
469	126.000000	6		0.008

FIGURE 6B

475	127.000000	3	0.004
478	128.000000	2	0.003
480	129.000000	4	0.005
484	130.000000	3	0.004
487	132.000000	3	0.004
490	133.000000	2	0.003
492	137.000000	1	0.001
493	140.000000	1	0.001
494	143.000000	1	0.001
495	152.000000	1	0.001
496	158.000000	1	0.001
497	166.000000	1	0.001
498	169.000000	2	0.003
500	170.000000	1	0.001
501	173.000000	2	0.003
503	176.000000	1	0.001
504	178.000000	1	0.001
505	180.000000	2	0.003
507	182.000000	1	0.001
508	185.000000	1	0.001
509	187.000000	1	0.001
510	190.000000	1	0.001
511	194.000000	1	0.001
512	195.000000	1	0.001
513	199.000000	1	0.001
514	206.000000	1	0.001
515	209.000000	1	0.001
516	216.000000	1	0.001
517	229.000000	1	0.001
518	230.000000	1	0.001
519	237.000000	1	0.001
520	239.000000	3	0.004
523	240.000000	2	0.003
525	241.000000	1	0.001
526	244.000000	2	0.003
528	247.000000	2	0.003
530	248.000000	2	0.003
532	250.000000	1	0.001
533	251.000000	1	0.001
534	254.000000	2	0.003
536	255.000000	2	0.003
538	256.000000	1	0.001
539	258.000000	1	0.001
540	272.000000	1	0.001
541	275.000000	1	0.001
542	280.000000	1	0.001
543	293.000000	1	0.001
544	318.000000	1	0.001
545	324.000000	1	0.001
546	340.000000	1	0.001
547	344.000000	1	0.001
548	346.000000	1	0.001
549	359.000000	1	0.001
550	365.000000	2	0.003
552	367.000000	1	0.001
553	377.000000	1	0.001
554	380.000000	1	0.001
555	387.000000	1	0.001
556	413.000000	1	0.001
557	418.000000	1	0.001
558	429.000000	1	0.001
559	453.000000	1	0.001
560	466.000000	1	0.001
561	472.000000	1	0.001
562	503.000000	2	0.003
564	504.000000	1	0.001
565	507.000000	1	0.001
566	510.000000	1	0.001
567	514.000000	1	0.001
568	517.000000	1	0.001
569	528.000000	1	0.001
570	570.000000	1	0.001
571	572.000000	1	0.001
572	586.000000	1	0.001
573	587.000000	1	0.001
574	601.000000	1	0.001
575	614.000000	1	0.001
576	624.000000	1	0.001
577	640.000000	1	0.001
578	663.000000	1	0.001
579	714.000000	1	0.001
580	716.000000	2	0.003
582	740.000000	1	0.001
583	742.000000	2	0.003
585	791.000000	1	0.001
586	853.000000	1	0.001
587	862.000000	1	0.001
588	866.000000	1	0.001
589	875.000000	1	0.001
590	885.000000	1	0.001
591	894.000000	1	0.001
592	926.000000	1	0.001
593	923.000000	1	0.001
594	941.000000	1	0.001

FIGURE 6C

595	976.000000	1	0.001
596	991.000000	1	0.001
597	1021.000000	1	0.001
598	1023.000000	1	0.001
599	1062.000000	1	0.001
600	1086.000000	2	0.003
602	1107.000000	1	0.001
603	1118.000000	1	0.001
604	1125.000000	1	0.001
605	1231.000000	1	0.001
606	1250.000000	1	0.001
607	1261.000000	2	0.003
609	1296.000000	1	0.001
610	1367.000000	1	0.001
611	1412.000000	1	0.001
612	1413.000000	1	0.001
613	1462.000000	1	0.001
614	1485.000000	1	0.001
615	1497.000000	1	0.001
616	1504.000000	1	0.001
617	1517.000000	1	0.001
618	1534.000000	1	0.001
619	1535.000000	2	0.003
621	1562.000000	1	0.001
622	1594.000000	1	0.001
623	1595.000000	1	0.001
624	1603.000000	1	0.001
625	1652.000000	1	0.001
626	1653.000000	1	0.001
627	1693.000000	1	0.001
628	1700.000000	1	0.001
629	1715.000000	1	0.001
630	1750.000000	1	0.001
631	1755.000000	1	0.001
632	1780.000000	1	0.001
633	1822.000000	1	0.001
634	1824.000000	1	0.001
635	1855.000000	1	0.001
636	1877.000000	1	0.001
637	1893.000000	1	0.001
638	1897.000000	1	0.001
639	1923.000000	1	0.001
640	1964.000000	1	0.001
641	1965.000000	1	0.001
642	2051.000000	1	0.001
643	2063.000000	1	0.001
644	2167.000000	1	0.001
645	2185.000000	2	0.003
647	2208.000000	1	0.001
648	2253.000000	1	0.001
649	2260.000000	1	0.001
650	2271.000000	1	0.001
651	2287.000000	1	0.001
652	2316.000000	1	0.001
653	2321.000000	1	0.001
654	2429.000000	1	0.001
655	2439.000000	1	0.001
656	2468.000000	1	0.001
657	2472.000000	1	0.001
658	2490.000000	1	0.001
659	2684.000000	1	0.001
660	2793.000000	1	0.001
661	2882.000000	1	0.001
662	2891.000000	1	0.001
663	2936.000000	1	0.001
664	3083.000000	1	0.001
665	3117.000000	1	0.001
666	3165.000000	1	0.001
667	3266.000000	1	0.001
668	3315.000000	1	0.001
669	3378.000000	1	0.001
670	3510.000000	1	0.001
671	3807.000000	1	0.001
672	3837.000000	1	0.001
673	3850.000000	1	0.001
674	3858.000000	1	0.001
675	3879.000000	1	0.001
676	3888.000000	1	0.001
677	3944.000000	1	0.001
678	3978.000000	1	0.001
679	4373.000000	1	0.001
680	4393.000000	1	0.001
681	4512.000000	1	0.001
682	4627.000000	1	0.001
683	4759.000000	1	0.001
684	4835.000000	1	0.001
685	5216.000000	1	0.001
686	5492.000000	1	0.001
687	5552.000000	1	0.001
688	5754.000000	1	0.001
689	6185.000000	1	0.001
690	6241.000000	1	0.001
691	6372.000000	1	0.001

FIGURE 6D

692	6385.000000	1	0.001
693	6659.000000	1	0.001
694	6816.000000	1	0.001
695	6821.000000	1	0.001
696	7307.000000	1	0.001
697	7329.000000	1	0.001
698	7746.000000	1	0.001
699	7927.000000	1	0.001
700	8039.000000	1	0.001
701	8053.000000	1	0.001
702	8353.000000	1	0.001
703	8847.000000	1	0.001
704	9206.000000	1	0.001
705	9256.000000	1	0.001
706	9517.000000	1	0.001
707	9541.000000	1	0.001
708	9562.000000	1	0.001
709	9965.000000	1	0.001
710	10020.000000	1	0.001
711	10376.000000	1	0.001
712	10837.000000	1	0.001
713	10918.000000	1	0.001
714	12042.000000	1	0.001
715	12793.000000	1	0.001
716	13012.000000	1	0.001
717	13179.000000	1	0.001
718	14692.000000	1	0.001
719	15825.000000	1	0.001
720	16877.000000	1	0.001
721	18885.000000	1	0.001
722	19273.000000	1	0.001
723	19696.000000	1	0.001
724	19849.000000	1	0.001
725	19851.000000	1	0.001
726	20840.000000	1	0.001
727	21440.000000	1	0.001
728	24573.000000	1	0.001
729	24770.000000	1	0.001
730	26278.000000	1	0.001
731	27238.000000	1	0.001
732	28133.000000	1	0.001
733	29513.000000	1	0.001
734	39867.000000	1	0.001
735	53138.000000	1	0.001
736	67271.000000	1	0.001

In addition, HISTLIST saved on printing time and paper. By printing the data in compressed form HISTLIST saved printing 448 lines (6 additional pages) in the case of telephone data 1 and 419 lines (5 additional pages) in the case of telephone data 2. Thus, HISTLIST not only gives the user more information than an ordered listing of the data, but also is cost effective in terms of printing time and paper used. Finally, note that it is not possible to look at the data in as much detail with routine HIST as with HISTLIST. If the data is continuous and there are no multiplicities, then HISTLIST gives only this information and an ordered listing of the data. The shape of the density function can best be seen (estimated) in using routine HIST.

IV. SECTIONING ROUTINE

A. DESCRIPTION

The third routine presented is the sectioning routine, HISTS. HISTS (sectioning routine) gives a way of assessing the variability of estimates of descriptive statistics from sample data. It is essential that the data be in random order.

The basic idea is as follows: Assume we have m independent observations $y_1, y_2, \ldots, y_m$ of a random variable Y. The usual estimate of its mean value $\mu = E(Y)$ is the sample mean $\bar{y}$, where $\bar{y} = \sum_{i=1}^{m} y_i/m$. Now $\bar{y}$ is the least-squares estimate of μ, and therefore unbiased with variance $\mathrm{var}(\bar{y}) = \sigma^2/m$, where $\sigma^2 = \mathrm{var}(y)$. Of course σ^2 is unknown, but we can estimate it from the data with the sample variance

$$s^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y})^2$$

and then estimate the variance of the estimate $\bar{y}$ of μ as

$$\widetilde{\mathrm{var}(\bar{y})} = \frac{s^2}{m} = \frac{1}{m(m-1)} \sum_{i=1}^{m} (y_i - \bar{y})^2$$

This is the basis for the sectioning routine: here the y_i are estimates of descriptive statistics from the m sections of the data and $\bar{y}$ is the average of the statistics

from each section. Estimates are assumed independent because the original data is assumed to be independent.

A complete description of how HISTS operates is contained in the variable HISTSHOW. When the user types HISTSHOW the following response is printed on the terminal:

```
HISTSHOW

SYNTAX  HISTS

   HISTS ALLOWS YOU TO INTERACTIVELY SECTION YOUR DATA AND
ASSESS THE VARIABILITY IN EACH OF THE DESCRIPTIVE STATISTICS
BY USING THE SECTIONED SAMPLE DATA.

   WHEN YOU TYPE HISTS YOU WILL BE ASKED TO DESIGNATE THE
NUMBER OF SECTIONS YOU DESIRE. HISTS WILL THEN TAKE
THE UNORDERED DATA AND DIVIDE THE DATA INTO THE NUMBER
OF SECTIONS YOU INDICATE DISCARDING ANY DATA POINTS LEFT
OVER. FOR EXAMPLE, IF YOU HAVE 301 DATA POINTS AND YOU
SELECT 10 SECTIONS HISTS WILL PLACE THE FIRST 30 DATA POINTS
IN THE FIRST SECTION, THE SECOND 30 DATA POINTS IN THE
SECOND SECTION AND SO ON UNTIL THE LAST DATA POINT IS
OMITTED. YOU WILL NOW HAVE 10 SECTIONS WITH 30 DATA POINTS
PER SECTION.

   HISTS WOULD NOW PRINT THE FOLLOWING STATISTICS ON EACH OF
THE SECTIONS: MEAN, MEDIAN, VARIANCE, STD DEV, COEF VAR,
SKEWNESS, KURTOSIS, MINIMUM AND MAXIMUM. IN ADDITION, THE
ABOVE STATISTICS WOULD BE PRINTED FOR THE UNSECTIONED DATA
TO ALLOW FOR COMPARISONS.

   FINALLY, HISTS WILL PRINT (1) THE MEAN OF THE SECTIONED
DATA STATISTICS. FOR EXAMPLE, THE MEAN FOR SKEWNESS WOULD BE
EACH SECTION VALUE FOR SKEWNESS SUMMED UP AND DIVIDED BY THE
NUMBER OF SECTIONS. (2) THE VARIANCE AND STD DEV OF THE
SECTIONED DATA STATISTICS. AND, (3) THE STD DEV DIVIDED BY
THE SQUARE ROOT OF THE NUMBER OF SECTIONS, WHICH ESTIMATES
THE STANDARD DEVIATION OF THE STATISTICS.

   AS A RESULT, HISTS WILL GIVE YOU AN UNBIASED ESTIMATE OF
THE VARIANCE OF THE SAMPLE MEAN, MEDIAN, VARIANCE, STD DEV,
COEF VAR, SKEWNESS AND KURTOSIS FROM USING THE SAMPLE
VARIANCE OF THE SECTIONED DATA. WITH THIS RESULT, CONFIDENCE
INTERVALS CAN ALSO BE OBTAINED FOR EACH OF THE ABOVE STATIS-
TICS, IF THE ESTIMATES FROM THE SECTIONS ARE NORMALLY DIS-
TRIBUTED. HISTS IS BEST SUITED FOR LARGE AND MODERATE SIZED
SAMPLES; FOR SMALL SAMPLES JACKNIFING SHOULD BE CONSIDERED.
```

B. USAGE WITH TELEPHONE DATA 1

HISTS was now used on telephone data 1 to assess the variability in the mean, median, variance, standard deviation, coefficient of variation, skewness and kurtosis. When HISTS was typed the following responses were entered (see figure 7).

The 672 data points of telephone data 1 were broken down into 16 sections with 42 data points per section. Because of this breakdown no data points were discarded.

The unsectioned statistics printed can be compared with the values printed by HIST (figure 1) and are in fact the same. Providing that the estimates are normally distributed (this can be checked with the normal plots, described later), confidence intervals for each of the statistics (mean, median, variance, standard deviation, coefficient of variation, skewness and kurtosis) based on the t-statistic can be obtained in the following manner

$$\bar{y}_n \pm \frac{s_{\bar{y}_n}}{\sqrt{m}} \; t_{(1-\frac{1}{2}\alpha),(m-1)}$$

Here $\bar{y}_n$ is the mean of the sectioned data statistics (obtained from column one under summary for sectioned data); $\frac{s_{\bar{y}_n}}{\sqrt{m}}$ is the standard deviation of the sectioned data statistic divided by the square root of the number of sections (obtained from column four under summary for sectioned data); m is the number sections chosen; and, $t_{(1-\frac{1}{2}\alpha),(m-1)}$ is the $1-\frac{1}{2}\alpha$ quantile of the t-distribution with m-1 degrees of freedom.

```
        HISTS
TYPE THE NUMBER OF SECTIONS YOU  DESIRE  ( INTEGER
BETWEEN 2 AND 28 )  BE SURE TO PICK YOUR NUMBER OF
SECTIONS SO AS TO  MINIMIZE  THE  NUMBER  OF  DATA
POINTS THAT WILL  HAVE  TO  BE  DISCARDED.  (HISTS
PLACES THE DATA INTO THE EQUAL NUMBER OF  SECTIONS
YOU  INDICATE  DISCARDING  ANY  DATA  LEFT  OVER )
[]:
        16

ENTER  YOUR  DATA  TO BE SECTIONED IN VECTOR  FORM
[]:
        TELDAT1
```

SECTION	MEAN	MEDIAN	VARIANCE	STD DEV	COFF VAR	SKEWNESS	KURTOSIS	MINIMUM	MAXIMUM
1	1.0526E03	8.5000E00	3.4598E07	5.8820E03	5.5879E00	6.3484E00	3.7831E01	1.0000E00	3.8003E04
2	3.2135E03	1.4500E01	1.8494E08	1.3599E04	4.2320E00	5.8106E00	3.3017E01	1.0000E00	8.5993E04
3	1.7662E03	1.4500E01	4.2383E07	6.5103E03	3.6860E00	4.2806E00	1.7830E01	1.0000E00	3.5644E04
4	6.0669E02	1.1000E01	5.3412E06	2.3111E03	3.8094E00	4.3148E00	1.6486E01	1.0000E00	1.1280E04
5	1.5639E03	5.0500E01	2.1924E07	4.6824E03	2.9941E00	4.2209E00	1.8565E01	1.0000E00	2.6443E04
6	2.5343E03	5.7000E01	4.0337E07	6.3511E03	2.5061E00	3.1573E00	9.5654E00	1.0000E00	3.0974E04
7	2.6778E03	2.2000E01	7.2756E07	8.5297E03	3.1853E00	4.1587E00	1.7579E01	1.0000E00	4.7120E04
8	9.8881E02	1.8500E01	3.8282E07	6.1873E03	6.2573E00	6.4801E00	3.8995E01	1.0000E00	4.0131E04
9	1.5176E03	2.2000E01	2.0792E07	4.5599E03	3.0046E00	2.9866E00	6.8551E00	1.0000E00	1.7174E04
10	2.7682E03	1.4000E01	1.0906E08	1.0443E04	3.7726E00	4.8932E00	2.4134E01	1.0000E00	6.1710E04
11	1.9258E03	1.4000E01	5.9852E07	7.7364E03	4.0173E00	5.4134E00	2.9059E01	1.0000E00	4.7592E04
12	8.1955E02	4.9500E01	8.2895E06	2.8791E03	3.5131E00	4.5999E00	1.9785E01	1.0000E00	1.5868E04
13	2.1201E03	4.0000E00	1.2224E08	1.1056E04	5.2150E00	5.9400E00	3.3861E01	1.0000E00	6.9775E04
14	2.3062E02	1.1500E01	3.3035E05	5.7476E02	2.4923E00	3.4695E00	1.2010E01	1.0000E00	2.9620E03
15	4.3752E02	7.0000E00	5.7201E06	2.3917E03	5.4664E00	6.3983E00	3.8289E01	1.0000E00	1.5504E04
16	5.4838E02	6.5000E00	1.1340E07	3.3675E03	6.1408E00	6.4765E00	3.8964E01	1.0000E00	2.1848E04
UNSECTIONED	1.5482E03	1.4000E01	4.8362E07	6.9543E03	4.4918E00	7.1531E00	6.2608E01	1.0000E00	8.5993E04

SUMMARY FOR SECTIONED DATA

	MEAN	VARIANCE	STD DEV	STD÷(SECS)*.5
MEAN	1.5482E03	8.6488E05	9.2999E02	2.3250E02
MEDIAN	2.0313E01	2.8023E02	1.6740E01	4.1850E00
VARIANCE	4.8637E07	2.6217E15	5.1203E07	1.2801E07
STD DEV	6.0664E03	1.2625E07	3.5532E03	8.8830E02
COEF VAR	4.1175E00	1.5503E00	1.2451E00	3.1128E-01
SKEWNESS	4.9343E00	1.4701E00	1.2125E00	3.0312E-01
KURTOSIS	2.4552E01	1.2484E02	1.1173E01	2.7933E00

FIGURE 7

C. INTERPRETATION OF RESULTS

As an example, a confidence interval for the coefficient of variation was obtained in the following manner. The mean value of the coefficient of variation for the 16 sections is 4.1175 (column 1). The standard deviation divided by the square root of 16 is .31128 (column 4). Using $\alpha = .05$, the t value with 15 degrees of freedom is 2.131. Thus, the 95% confidence interval for the coefficient of variation for telephone data 1 is $4.1175 \pm (.31128)(2.131)$ which is [3.454, 4.781]. Confidence intervals on the six other statistics could be obtained in the same fashion.

Again note that the use of the variance estimate from the sectioned data to give confidence intervals is based on the assumption that the estimates from the sections are independent and normally distributed. The normality will depend on the number of observations in each section, which should be kept large to induce normality. This requirement conflicts with the need to make the number of sections large to reduce the variability in the estimate of the variance of the statistics.

Another problem is that if the number of observations in each section is small, the estimates may be severely biased. This effect can be seen in figure 7: note that all of the 16 estimates of skewness from the sections are smaller than the estimate 7.1531 from the unsectioned data.

V. JACKNIFE ROUTINE

A. DESCRIPTION

The fourth routine presented is the jacknife routine. HISTJACK (jacknife routine) is another way of assessing the variability in the estimates from sample data, and also of reducing bias in estimates of the descriptive statistics.

The jacknife procedure, like the previous sectioning method, is based on the assumption that an independent and identically distributed random sample $x_1, x_2, \ldots, x_n$ have come from a population with an unknown distribution function $F_X(x)$. If we divide the sample into r groups, with each group containing the same number of elements, we can obtain estimates $\tilde{\theta}$ of the descriptive statistics, which we denote generically as θ, in the same manner as previously done with the sectioning method. The difference here is that the descriptive statistics are computed with the j^{th} group deleted $j=1,2,\ldots,r$. We then let $\tilde{\theta}_{(j)}$ be the result or the descriptive statistic estimate computed with the j^{th} subgroup omitted, and $\tilde{\theta}_{all}$ is the corresponding result or descriptive statistic estimated from the entire sample (no group omitted). The jacknife pseudo-values are then computed in the following way:

$$\tilde{\theta}_{*j} = (r)(\tilde{\theta}_{all}) - (r-1)(\tilde{\theta}_{(j)}) \qquad j = 1,2,\ldots,r$$

Then we define the jacknifed estimator to be:

$$\tilde{\theta}_* = \frac{1}{r} \sum_{j=1}^{r} \tilde{\theta}_{*j}$$

The pseudo-values can be used to obtain variance estimates for $\tilde{\theta}_*$, and to set approximate confidence limits, using Student's t. The idea is that the pseudo-values will be approximately independent and possibly normally distributed. The jacknifed estimator $\tilde{\theta}_*$ is a sample average so we form an estimate $s_*{}^2$ of its variance given by the following relationship (Miller, 1974):

$$s^2 = \frac{\Sigma\tilde{\theta}^2_{*j} - \frac{1}{r}(\Sigma\tilde{\theta}_{*j})^2}{r-1}$$

$$s^2_* = s^2/r$$

This procedure is particularly useful if the number n of data points is small, but it must be used with care. Note, that the estimator $\tilde{\theta}_*$ is designed to eliminate a $1/n$ bias term in the estimator $\tilde{\theta}$.

A complete description of how HISTJACK operates is contained in the variable HISTJACKHOW. When the user types HISTJACKHOW the following response is printed on the terminal.

HISTJACKHOW

SYNTAX HISTJACK

HISTJACK ALLOWS YOU TO INTERACTIVELY JACKNIFE YOUR DATA AND ASSESS THE VARIABILITY IN EACH OF THE STATISTICAL ESTIMATES BY USING THE SAMPLE DATA.

WHEN YOU TYPE HISTJACK YOU WILL BE ASKED TO DESIGNATE THE NUMBER OF GROUPS YOU DESIRE. HISTJACK WILL TAKE THE UNORDERED DATA AND DIVIDE THE DATA INTO THE NUMBER OF GROUPS YOU INDICATE DISCARDING ANY DATA POINTS LEFT OVER. FOR EXAMPLE, IF YOU HAVE 22 DATA POINTS AND YOU SELECT 7 GROUPS HISTJACK WILL PLACE THE FIRST 3 DATA POINTS IN GROUP 1, THE SECOND 3 DATA POINTS IN GROUP 2, AND SO ON UNTIL THE LAST DATA POINT IS OMITTED. YOU WOULD NOW HAVE 7 GROUPS WITH 3 DATA POINTS PER GROUP. IF YOU HAD ELECTED TO DO A COMPLETE JACKNIFE, THAT IS TYPED 22, YOU WOULD NOW HAVE 22 GROUPS WITH 1 DATA POINT OMITTED PER GROUP.

HISTJACK WOULD NOW PERFORM STATISTICAL COMPUTATIONS USING THE JACKNIFE PROCEDURE. THAT IS, BY OMITTING ONE GROUP AT A TIME, STARTING WITH THE FIRST GROUP, HISTJACK WOULD PRINT THE FOLLOWING STATISTICS: MEAN, MEDIAN, VARIANCE, STD DEV, COEF VAR, SKEWNESS, KURTOSIS, MINIMUM AND MAXIMUM. IN ADDITION, THE ABOVE STATISTICS WOULD BE PRINTED FOR THE UNGROUPED DATA TO ALLOW FOR COMPARISONS. (NOTE, THE COLUMNS GIVE THE STATISTIC ESTIMATED FROM ALL THE DATA WITH ONE GROUP MISSING, AND NOT THE PSEUDO-VALUES)

FINALLY, HISTJACK WILL PRINT (1) THE JACKNIFE ESTIMATE (2) THE SAMPLE VARIANCE OF THE PSEUDO-VALUES DERIVED IN THE JACKNIFE ESTIMATE (3) AND, THE ESTIMATED STD DEV OF THE JACKNIFE ESTIMATE DIVIDED BY THE SQUARE ROOT OF THE NUMBER OF GROUPS.

AS A RESULT, HISTJACK WILL GIVE YOU AN ESTIMATE OF THE VARIANCE OF THE SAMPLE MEAN, MEDIAN, VARIANCE, STD DEV, COEF VAR, SKEWNESS AND KURTOSIS USING THE SAMPLE VARIANCE OF THE JACKNIFED DATA. WITH THIS RESULT, CONFIDENCE INTERVALS CAN BE OBTAINED FOR EACH OF THE ABOVE STATISTICS, AGAIN ASSUMING THAT THE PSEUDO-VALUES ARE APPROXIMATELY INDEPENDENT AND NORMALLY DISTRIBUTED. HISTJACK IS BEST SUITED FOR SMALL SAMPLES.

B. USAGE WITH TELEPHONE DATA 1

HISTJACK was now used on telephone data 1 to assess the variability in the mean, median, variance, standard deviation, coefficient of variation, skewness and kurtosis. When HISTJACK was typed the following responses were entered. (see figure 8)

The 672 data points were broken down into 16 groups with 42 data points per group. Again, because of this breakdown no data points were discarded.

The ungrouped statistics printed are again the same values that were printed by HIST (figure 1). Using the jacknife method, confidence intervals for each of the statistics (mean, median, variance, standard deviation, coefficient of variation, skewness and kurtosis) can be obtained in the following manner;

$$\tilde{\theta}_* \pm (s_*)\, t_{(1-\frac{1}{2}\alpha),(r-1)} \cdot$$

Here $\tilde{\theta}_*$ is the jacknife estimate of the sample data (obtained from column one under summary for jacknifed data); s_* is the jacknife estimate of the standard deviation divided by the square root of the number of groups (obtained from column four under summary for jacknifed data); r is the number of groups chosen; and, $t_{(1-\frac{1}{2}\alpha),(r-1)}$ is the $1-\frac{1}{2}\alpha$ quantile of the t-distribution with r-1 degrees of freedom. The basis for these assertions about the confidence intervals using the jacknifing technique is asymptotic and great care must be taken in using them.

FIGURE 8

HISTJACK
TYPE THE NUMBER OF GROUPS YOU DESIRE (INTEGER BETWEEN 2 AND 50) BE SURE TO PICK YOUR NUMBER OF GROUPS SO AS TO MINIMIZE THE NUMBER OF DATA POINTS THAT WILL HAVE TO BE DISCARDED. (HISTJACK PLACES THE DATA INTO THE EQUAL NUMBER OF GROUPS YOU INDICATE DISCARDING ANY DATA LEFT OVER)
[]:
16

ENTER YOUR DATA TO BE JACKNIFED IN VECTOR FORM
[]:
TELDAT1

GROUP	MEAN	MEDIAN	VARIANCE	STD DEV	COEF VAR	SKEWNESS	KURTOSIS	MINIMUM	MAXIMUM
1	1.5813E03	1.5000E01	4.9318E07	7.0227E03	4.4412E00	7.1746E00	6.3025E01	1.0000E00	8.5993E04
2	1.4372E03	1.4000E01	3.9338E07	6.2720E03	4.3641E00	6.5220E00	5.0690E01	1.0000E00	6.9775E04
3	1.5337E03	1.4000E01	4.8825E07	6.9875E03	4.5560E00	7.3093E00	6.4762E01	1.0000E00	8.5993E04
4	1.6110E03	1.4000E01	5.1180E07	7.1540E03	4.4408E00	6.9781E00	5.9257E01	1.0000E00	8.5993E04
5	1.5472E03	1.3500E01	5.0162E07	7.0825E03	4.5777E00	7.1494E00	6.1827E01	1.0000E00	8.5993E04
6	1.4825E03	1.3000E01	4.8893E07	6.9923E03	4.7166E00	7.3663E00	6.5160E01	1.0000E00	8.5993E04
7	1.4729E03	1.3000E01	4.6758E07	6.8380E03	4.6425E00	7.5081E00	6.8831E01	1.0000E00	8.5993E04
8	1.5855E03	1.4000E01	4.9073E07	7.0052E03	4.4183E00	7.1850E00	6.3371E01	1.0000E00	8.5993E04
9	1.5503E03	1.3500E01	5.0236E07	7.0877E03	4.5720E00	7.1592E00	6.1773E01	1.0000E00	8.5993E04
10	1.4669E03	1.4000E01	4.4376E07	6.6615E03	4.5413E00	7.4572E00	6.9618E01	1.0000E00	8.5993E04
11	1.5230E03	1.4000E01	4.7680E07	6.9050E03	4.5337E00	7.3200E00	6.5949E01	1.0000E00	8.5993E04
12	1.5968E03	1.3000E01	5.1013E07	7.1423E03	4.4729E00	7.0092E00	5.9701E01	1.0000E00	8.5993E04
13	1.5101E03	1.6000E01	4.3600E07	6.6030E03	4.3726E00	7.1493E00	6.5032E01	1.0000E00	8.5993E04
14	1.6361E03	1.4000E01	5.1446E07	7.1726E03	4.3841E00	6.9200E00	5.8526E01	1.0000E00	8.5993E04
15	1.6223E03	1.5000E01	5.1130E07	7.1506E03	4.4078E00	6.9784E00	5.9319E01	1.0000E00	8.5993E04
16	1.6149E03	1.5000E01	5.0781E07	7.1261E03	4.4128E00	7.0291E00	6.0129E01	1.0000E00	8.5993E04
UNGROUPED	1.5482E03	1.4000E01	4.8362E07	6.9543E03	4.4918E00	7.1531E00	6.2608E01	1.0000E00	8.5993E04

SUMMARY FOR JACKNIFED DATA

	JACKNIFE ESTIMATE	VARIANCE	(VAR÷GROUPS)*.5 JACKNIFE ESTIMATE OF STD DEV OF MEAN OF PSEUDO-VALUES
MEAN	1.5482E03	8.6488E05	2.3250E02
MEDIAN	1.3063E01	1.5656E02	3.1281E00
VARIANCE	4.8344E07	2.5453E15	1.2613E07
STD DEV	7.0154E03	1.3879E07	9.3135E02
COEF VAR	4.5053E00	2.4262E00	3.8940E-01
SKEWNESS	7.3732E00	1.2963E01	9.0012E-01
KURTOSIS	6.7077E01	4.6806E03	1.7104E01

C. INTERPRETATION OF RESULTS

To compare the confidence interval obtained for the coefficient of variation using the sectioning routine with that obtained using the jacknife routine the following was done. The jacknife estimate of the coefficient of variation for the 16 groups is 4.5053 (column 1). The jacknife estimate of the standard deviation divided by the square root of 16 is .3894 . Using $\alpha = .05$, the t value with 15 degrees of freedom is 2.131. Thus, the 95% confidence interval for the coefficient of variation for telephone data 1 is 4.5053 $\pm$ (.3894) (2.131) which is [3.676, 5.335]. This compares with the confidence interval of [3.454, 4.781] using the sectioning routine described in section IV. Likewise, confidence intervals on the remaining six statistics could be obtained in a similar manner. Note that the values obtained for the skewness coefficient from the sections are now not evidently biased; of the 16 values, 7 have values below the value 7.1531 for all the data.

D. USAGE WITH COST OVERRUN DATA

To demonstrate how the complete jacknife could be used and why it is better to use when possible, the following was done. The 22 data points of the cost overrun data were used with the jacknife routine (HISTJACK). When HISTJACK was typed the data was entered in the variable YROVR and 22 was typed as the number of groups. By typing 22, which is the same as the number of data points, a complete jacknife was done.

Looking at the output from the complete jacknife (figure 9), the cost overrun data can be studied. One can note that by using the complete jacknife the mean, median, and variance of the jacknife estimate (column one under summary for jack-nifed data) are the same value as the ungrouped mean, median and variance. But, also note that the coefficient of variation is less than zero which can happen when using the jack-nife technique.

```
      HISTJACK
TYPE THE NUMBER  OF  GROUPS   OU  DESIRE  (INTEGER
BETWEEN  2  AND  50 )  BE SURE TO PICK YOUR NUMBER
OF GROUPS SO AS TO MINIMIZE  THE  NUMBER  OF  DATA
POINTS THAT WILL HAVE TO BE DISCARDED.   (HISTJACK
PLACES THE DATA INTO THE EQUAL  NUMBER  OF  GROUPS
YOU INDICATE DISCARDING ANY DATA LEFT OVER)
[]:
      22

ENTER YOUR DATA  TO  BE  JACKNIFED  IN VECTOR FORM
[]:
      YROVR
```

GROUP	MEAN	MEDIAN	VARIANCE	STD DEV	COEF VAR	SKEWNESS	KURTOSIS	MINIMUM	MAXIMUM
1	1.0524E00	-1.6000E00	1.0228E02	1.0113E01	9.6101E00	7.7349E-01	7.3991E-02	-1.3600E01	2.5300E01
2	1.2048E00	-1.2000E00	1.0189E02	1.0094E01	8.3784E00	7.2904E-01	5.1529E-02	-1.3600E01	2.5300E01
3	1.3190E00	-1.2000E00	1.0089E02	1.0044E01	7.6149E00	7.0897E-01	7.9345E-02	-1.3600E01	2.5300E01
4	1.7714E00	-1.2000E00	9.1014E01	9.5401E00	5.3856E00	8.7130E-01	3.2139E-01	-1.3000E01	2.5300E01
5	9.9048E-01	-1.6000E00	1.0213E02	1.0106E01	1.0203E01	7.9519E-01	1.0417E-01	-1.3600E01	2.5300E01
6	7.6190E-01	-1.6000E00	1.0006E02	1.0003E01	1.3129E01	8.8022E-01	3.1867E-01	-1.3600E01	2.5300E01
7	1.2286E00	-1.2000E00	1.0173E02	1.0086E01	8.2096E00	7.2372E-01	5.4384E-02	-1.3600E01	2.5300E01
8	1.4381E00	-1.2000E00	9.9207E01	9.9603E00	6.9260E00	7.0632E-01	1.4054E-01	-1.3600E01	2.5300E01
9	1.5286E00	-1.2000E00	9.7491E01	9.8738E00	6.4595E00	7.2120E-01	2.0046E-01	-1.3600E01	2.5300E01
10	1.2000E00	-1.2000E00	1.0192E02	1.0095E01	8.4128E00	7.3016E-01	5.1155E-02	-1.3600E01	2.5300E01
11	1.0190E00	-1.6000E00	1.0222E02	1.0111E01	9.9216E00	7.8499E-01	8.8707E-02	-1.3600E01	2.5300E01
12	1.7429E00	-1.2000E00	9.1918E01	9.5874E00	5.5009E00	8.4272E-01	3.1785E-01	-1.3600E01	2.5300E01
13	1.6000E00	-1.2000E00	9.5869E01	9.7913E00	6.1195E00	7.4622E-01	2.4899E-01	-1.3600E01	2.5300E01
14	1.2857E00	-1.2000E00	1.0124E02	1.0062E01	7.8260E00	7.1331E-01	6.7672E-02	-1.3600E01	2.5300E01
15	6.0476E-01	-1.6000E00	9.7232E01	9.8607E00	1.6305E01	9.3010E-01	5.4286E-01	-1.3600E01	2.5300E01
16	1.9048E-01	-1.6000E00	8.4311E01	9.1821E00	4.8206E01	8.9026E-01	9.9254E-01	-1.3600E01	2.5300E01
17	6.8571E-01	-1.6000E00	9.8831E01	9.9414E00	1.4498E01	9.0629E-01	4.2160E-01	-1.3600E01	2.5300E01
18	-8.0952E-02	-1.6000E00	7.1546E01	8.4585E00	1.0449E02	4.6734E-01	-2.0707E-01	-1.3600E01	1.9600E01
19	1.1810E00	-1.6000E00	1.0202E02	1.0101E01	8.5529E00	7.3487E-01	5.0328E-02	-1.3600E01	2.5300E01
20	9.6190E-01	-1.6000E00	1.0201E02	1.0100E01	1.0500E01	8.0563E-01	1.2225E-01	-1.3600E01	2.5300E01
21	1.3333E00	-1.2000E00	1.0072E02	1.0036E01	7.5270E00	7.0754E-01	8.5165E-02	-1.3600E01	2.5300E01
22	5.8095E-01	-1.6000E00	9.6705E01	9.8339E00	1.6927E01	9.3606E-01	5.8006E-01	-1.3600E01	2.5300E01
UNGROUPED	1.0727E00	-1.4000E00	9.7420E01	9.8702E00	9.7010E00	7.8191E-01	2.2400E-01	-1.3600E01	2.5300E01

SUMMARY FOR JACKNIFED DATA

	JACKNIFE ESTIMATE	VARIANCE	(VAR÷GROUPS)*.5 JACKNIFE ESTIMATE OF STD DEV OF MEAN OF PSEUDO-VALUES
MEAN	1.0727E00	9.7420E01	2.1043E00
MEDIAN	-1.4000E00	1.8480E01	9.1652E-01
VARIANCE	9.7420E01	2.3840E04	3.2919E01
STD DEV	1.0026E01	6.7477E01	1.7513E00
COEF VAR	-1.2279E02	2.0893E05	9.7452E01
SKEWNESS	8.7458E-01	4.8496E00	4.6951E-01
KURTOSIS	4.3524E-01	2.7843E01	1.1250E00

FIGURE 9

VI. EXPONENTIAL PLOTTING ROUTINE

A. DESCRIPTION

The fifth routine presented is an exponential plotting routine. Routine EXPONP is a way of plotting the data to see if it "fits" an exponential distribution, and also to give some indication of what alternative distributions could be used if the exponential hypothesis is rejected.

A complete description of how EXPONP operates is contained in the variable EXPONPHOW . When the user types EXPONPHOW the following response is printed on the terminal.

```
      EXPONPHOW

SYNTAX  EXPONP

   EXPONP  ORDERS  THE  DATA  X(I)  AND   COMPUTES
THE EMPIRICAL LOG SURVIVER FUNCTION FOR THE  DATA.
THAT IS,

            \ /          |    |\ |   /        I   \
             X       VS  |    | \|  | 1 -   -----  |
            / \ (I)      |__  |  |   \       N+1  /

THE ORDERED DATA IS PLOTTED AGAINST THE  LOG  SUR-
VIVER FUNCTION TO SEE IF THERE IS  A  LINEAR  FIT.
EXPONP ALSO ALLOWS YOU TO TITLE YOUR PLOT.
```

B. USAGE WITH TELEPHONE DATA 1

EXPONP was used with telephone data 1 to see if the data plotted as a relative straight line. When EXPONP was typed the following responses were entered.

```
EXPONP
EXPONP ORDERS THE DATA YOU GIVE AND COMPUTES THE
EMPIRICAL  LOG  SURVIVER  FUNCTION FOR THE DATA.
A PLOT OF THE LOG SURVIVER FUNCTION FOR THE DATA
IS THEN PRINTED TO SEE IF THERE IS A LINEAR FIT.

IF YOU WANT TO TITLE  YOUR PLOT TYPE YOUR TITLE.
IF YOU DO NOT WANT A TITLE JUST HIT THE CARRIAGE
RETURN.

TELEPHONE DATA 1

ENTER YOUR DATA IN VECTOR FORM
⎕:
      TELDAT1
```

Looking at figure 10 (plot of telephone data 1 using EXPONP), it was found that the data did not plot linearly from the origin, but that the data did appear somewhat linear in the tail (5,000 to 90,000 range).

C. USAGE WITH RANDOM GENERATED EXPONENTIALLY DISTRIBUTED SAMPLE WITH MEAN SAME AS TELEPHONE DATA 1

As a comparison, EXPONP was used with an exponentially generated random sample with the same mean as telephone data 1 (figure 11). As expected, this plot is, within limits of sample fluctuations, linear from the origin and in fact, what telephone data 1 would have looked like if the data was truly exponential. The quantization because of the coarseness of the APL type-ball is evident in this plot. The sample size is 672 , but not all these points can be plotted separately.

FIGURE 10

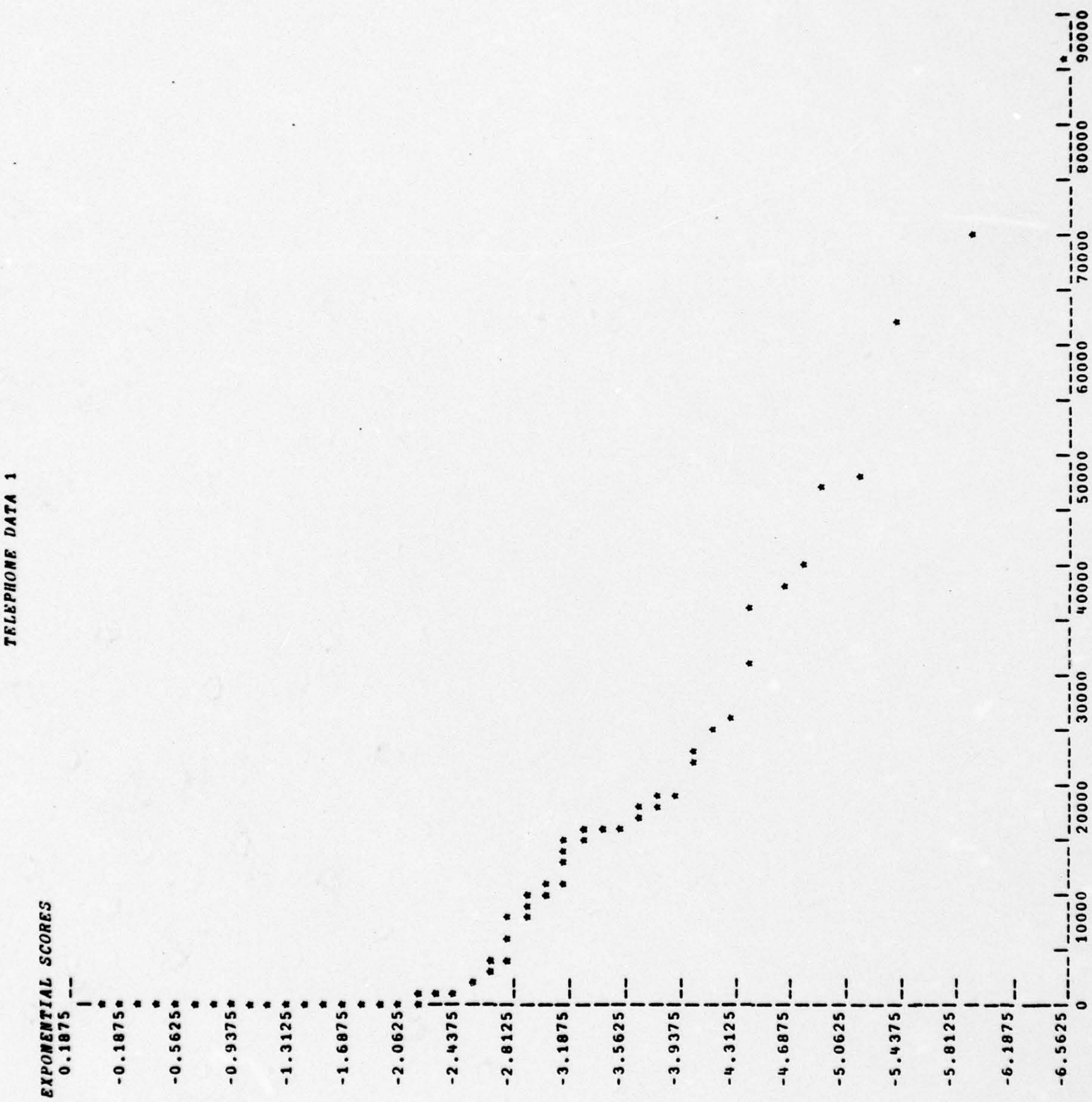
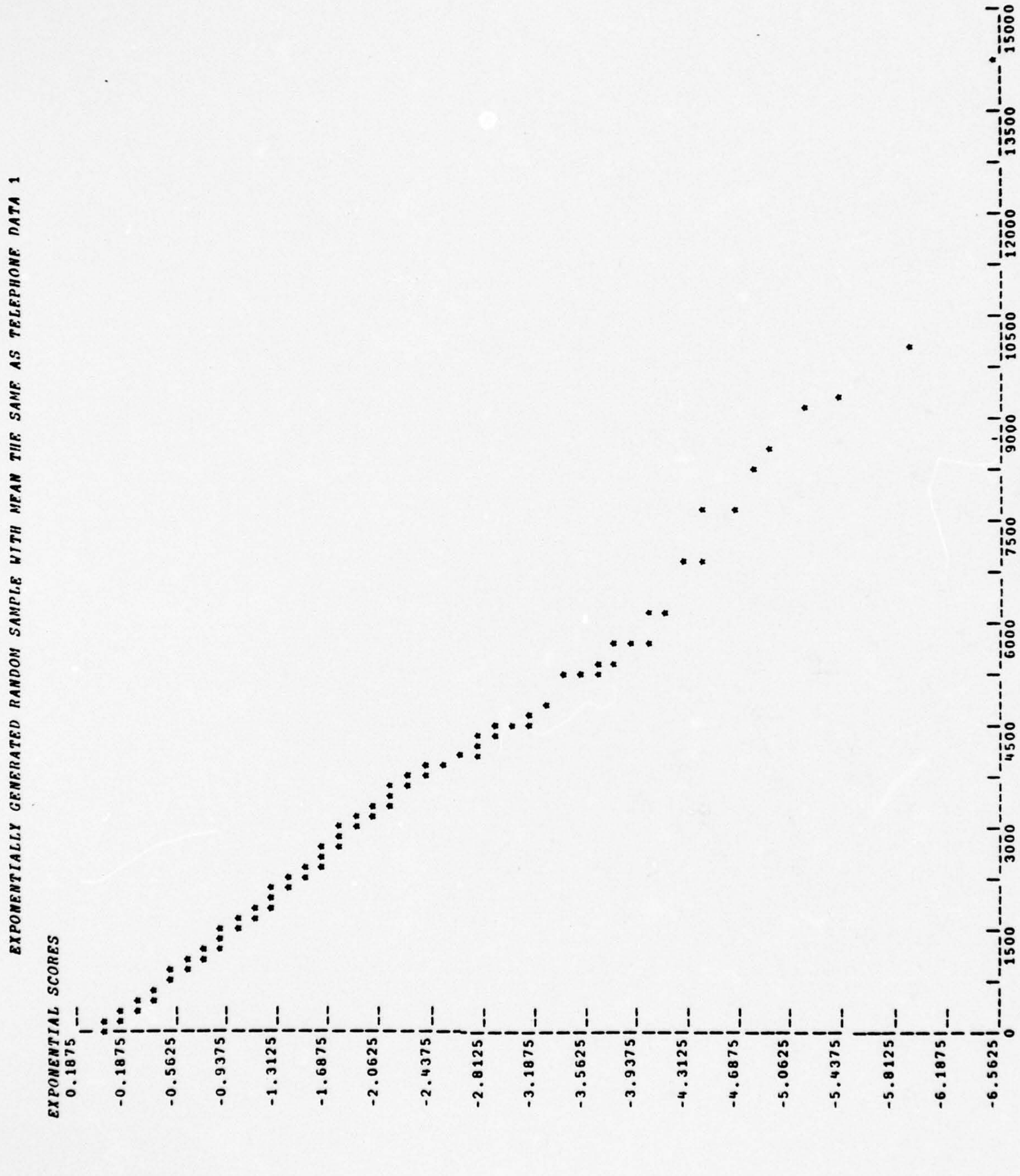

TELEPHONE DATA 1
EXPONENTIAL SCORES
0.1875
-0.1875
-0.5625
-0.9375
-1.3125
-1.6875
-2.0625
-2.4375
-2.8125
-3.1875
-3.5625
-3.9375
-4.3125
-4.6875
-5.0625
-5.4375
-5.8125
-6.1875
-6.5625
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
ORDERED DATA

FIGURE 11


EXPONENTIALLY GENERATED RANDOM SAMPLE WITH MEAN THE SAME AS TELEPHONE DATA 1
EXPONENTIAL SCORES
0.1875
-0.1875
-0.5625
-0.9375
-1.3125
-1.6875
-2.0625
-2.4375
-2.8125
-3.1875
-3.5625
-3.9375
-4.3125
-4.6875
-5.0625
-5.4375
-5.8125
-6.1875
-6.5625
0
1500
3000
4500
6000
7500
9000
10500
12000
13500
15000
ORDERED DATA

VII. NORMAL PLOTTING ROUTINE

A. DESCRIPTION

The final routine presented is a normal plotting routine. Routine NORMP is a way of plotting the data to see if it "fits" a normal distribution. In particular one might want to look at estimates of descriptive statistics obtained from sections and groups in routines HISTS and HISTJACK .

A complete description of how NORMP operates is contained in the variable NORMPHOW . When the user types NORMPHOW the following response is printed on the terminal.

NORMPHOW

SYNTAX NORMP

NORMP ORDERS THE DATA X(I) AND COMPUTES THE INVERSE OF THE UNIT NORMAL CUMULATIVE DISTRIBUTION. THAT IS,

$$X_{(I)} \quad \text{VS} \quad \Phi^{-1}\left(\frac{I}{N+1}\right)$$

THE ORDERED DATA IS PLOTTED AGAINST THE INVERSE OF THE UNIT NORMAL CUMULATIVE DISTRIBUTION TO SEE IF THERE IS A LINEAR FIT. NORMP ALSO ALLOWS YOU TO CONVIENTLY TITLE YOUR PLOT.

B. USAGE WITH COST OVERRUN DATA

NORMP was used with the cost overrun data to see if the data plotted as a relative straight line. When NORMP was typed the following responses were entered.

```
      NORMP
NORMP ORDERS THE DATA YOU GIVE  AND  COMPUTES THE
INVERSE OF THE UNIT NORMAL  CUMULATIVE  DISTRIBU-
TION FOR THE DATA.  A PLOT OF THE INVERSE  OF THE
UNIT NORMAL CUMULATIVE DISTRIBUTION VS THE ORDER-
ED DATA IS THEN  PRINTED  TO  SEE  IF  THERE IS A
LINEAR FIT.

IF YOU  WANT TO TITLE YOUR PLOT TYPE YOUR  TITLE.
IF YOU DO NOT WANT A TITLE JUST HIT  THE CARRIAGE
RETURN.

COST OVERRUNS

ENTER YOUR DATA IN VECTOR FORM
⎕:
      YROVR
```

Note that the cost overrun data was contained in the variable YROVR . Looking at figure 12 (plot of cost overrun data using NORMP), it was found that the data did in fact plot fairly linear through the range -14 to 26 (formal tests are available; see Wilk & Gnanadesikan, 1968).

C. USAGE WITH NORMAL SAMPLE GENERATED WITH MEAN AND VARIANCE THE SAME AS COST OVERRUN DATA

As a comparison, NORMP was used with a normal sample with the same mean and variance as the cost overrun data (figure 13). As expected, this plot is very linear. But again, this plot is not that much different from that of figure 12, which gives credence to the fact that the cost overrun data might in fact be normally distributed.

FIGURE 12

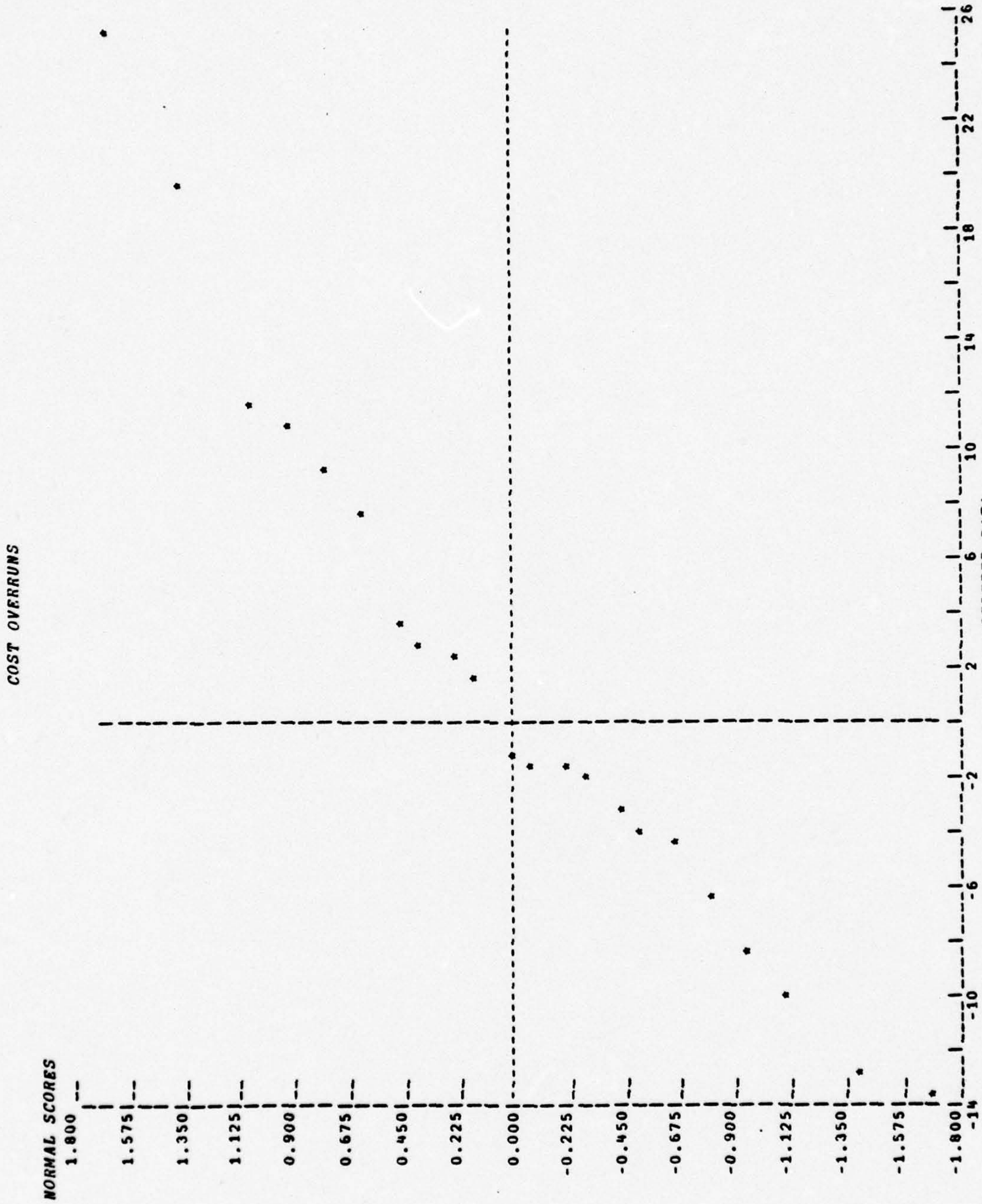
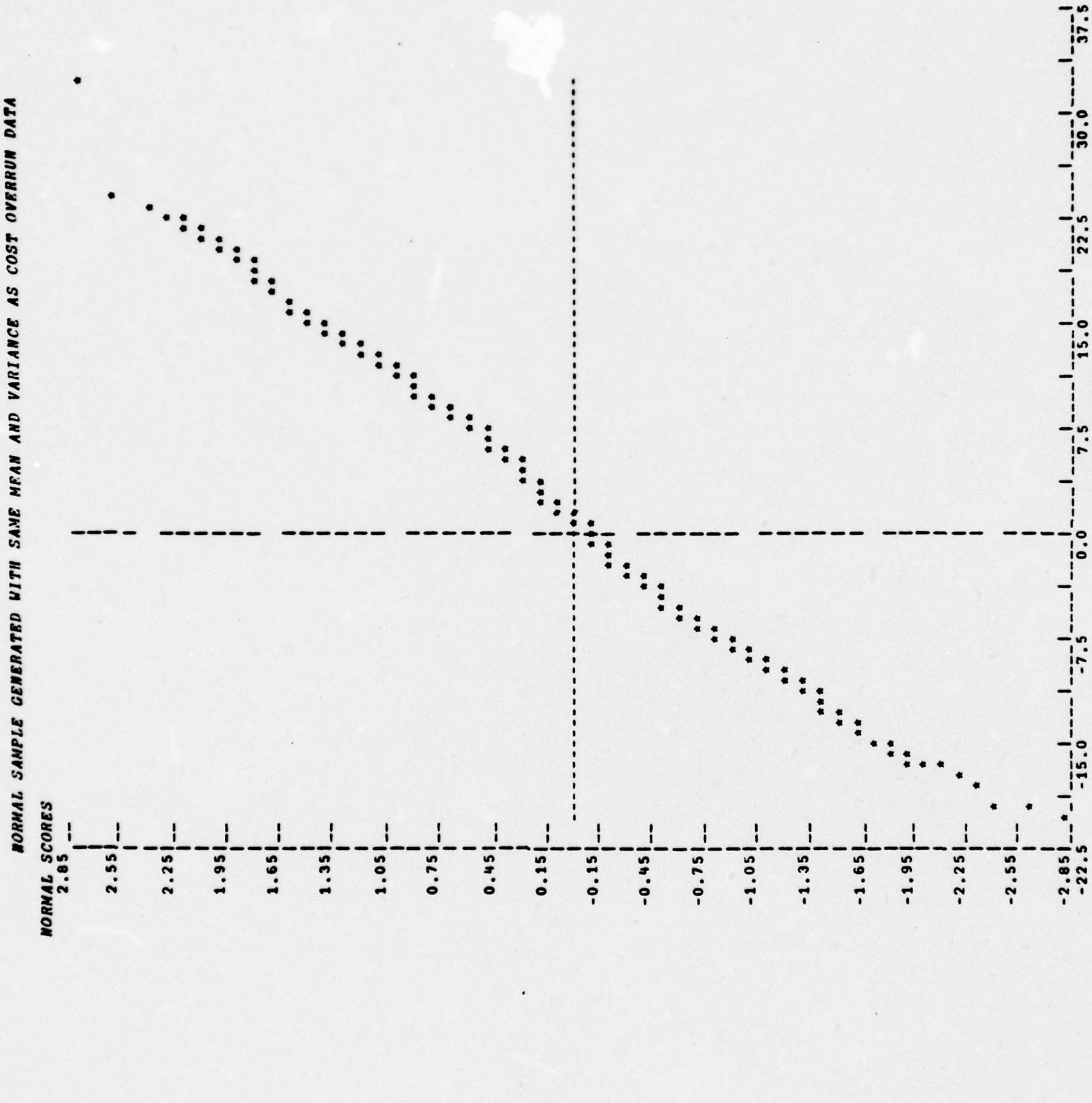

COST OVERRUNS
NORMAL SCORES
ORDERED DATA
1.800
1.575
1.350
1.125
0.900
0.675
0.450
0.225
0.000
-0.225
-0.450
-0.675
-0.900
-1.125
-1.350
-1.575
-1.800
-14
-10
-6
-2
2
6
10
14
18
22
26

FIGURE 13

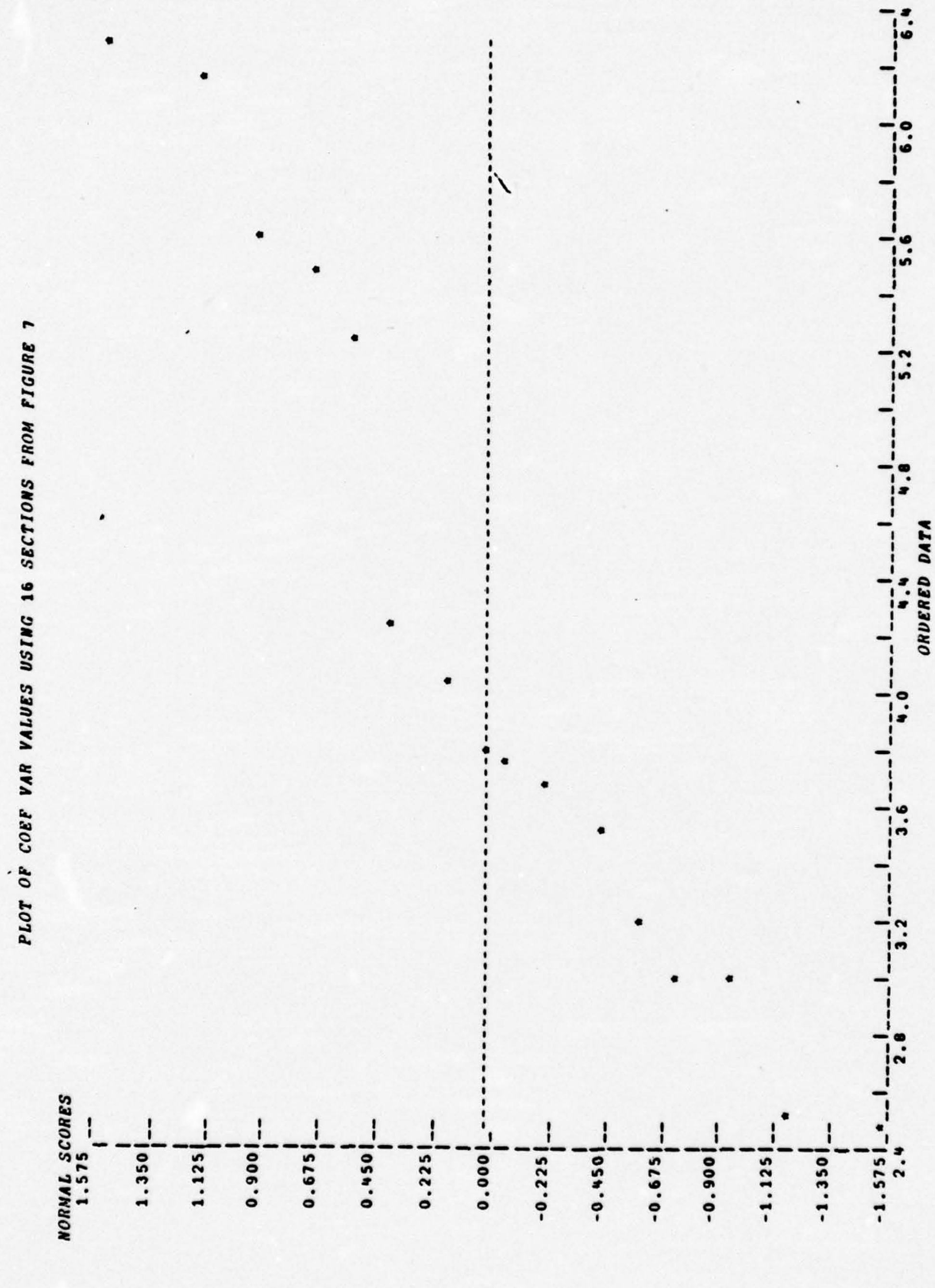


NORMAL SAMPLE GENERATED WITH SAME MEAN AND VARIANCE AS COST OVERRUN DATA
NORMAL SCORES
2.85
2.55
2.25
1.95
1.65
1.35
1.05
0.75
0.45
0.15
-0.15
-0.45
-0.75
-1.05
-1.35
-1.65
-1.95
-2.25
-2.55
-2.85
-22.5
-15.0
-7.5
0.0
7.5
15.0
22.5
30.0
37.5
ORDERED DATA

D. USAGE WTH COEFFICIENT OF VARIATION DATA OBTAINED FROM USING SECTIONING ROUTINE

In order to check for normality in the sectioned estimates obtained from using HISTS (sectioning routine) the following was done. The 16 coefficient of variation values obtained from using HISTS with telephone data 1 (column 5, figure 7) were entered as a vector into NORMP. Figure 14 shows that the plot is marginally linear. This demonstrates the need for formal tests to verify normality in the absence of a strictly linear plot (Wilk & Gnanadsikan, 1968).

FIGURE 14


PLOT OF COEF VAR VALUES USING 16 SECTIONS FROM FIGURE 7
NORMAL SCORES
1.575
1.350
1.125
0.900
0.675
0.450
0.225
0.000
-0.225
-0.450
-0.675
-0.900
-1.125
-1.350
-1.575
2.4
2.8
3.2
3.6
4.0
4.4
4.8
5.2
5.6
6.0
6.4
ORDERED DATA

VIII. THE INDEPENDENCE AND MARKOV CHAIN HYPOTHESES FOR THE TELEPHONE DATA

The telephone data used in the thesis (Lewis & Cox, 1966) actually consists of binary bits transmitted over telephone lines and the information that the bit transmitted at time i, $i = 0,1,2,\ldots$ is in error or not. This information is characterized by a sequence of binary-valued random variables $x(i)$, $i = 0,1,\ldots$ where $x(i)=1$ means that the bit transmitted at time i is in error, while $x(i)=0$ means that the bit transmitted at time zero is correctly transmitted.

In telephone data 1 there are 672 ones and 1,105,476 zeros, and a much more compact and equivalent representation of the data is obtained via the sequence of random variables $y(j)$, $j=1,2,\ldots$ where $y(j)$ is one plus the number of correctly transmitted bits between the j^{th} and $(j-1)^{st}$ bit error, with the convention that $y(j)=1$ if the errors occur on adjacent transmitted bits, and $y(1)$ is the time from $i=0$ to the first incorrectly transmitted bit. The $y(j)$ are called the times-between-errors.

A *null hypothesis* for the error structure which could be examined is that errors occur independently at each bit with a fixed probability, i.e.

$$P\{x(i)=1\} = \pi(1) \qquad i=0,1,\ldots$$

$$P\{x(i)=0\} = \pi(0) = 1-\pi(1) \qquad i=0,1,\ldots$$

The $y(j)$'s then are independent and geometrically distributed, since

$$P\{y(j)=1\} = P\{\text{if } (j-1)^{st} \text{ error at time } i;\ j^{th} \text{ at time } i+1\}$$

$$= \pi(1)$$

$$P\{y(j)=2\} = P\{\text{if } (j-1)^{st} \text{ error at time } i;\ j^{th} \text{ at time } i+2\}$$

$$= \pi(1)[1-\pi(1)] = \pi(1)\pi(0)$$

$$P\{y(j)=k+1\} = P\{\text{if } (j-1)^{st} \text{ error at time } i;\ j^{th} \text{ at time } i+1+k\}$$

$$= \pi(1)[1-\pi(1)]^k = \pi(1)[\pi(0)]^k$$

Note that, using the geometric series summation formula,

$$\sum_{k=1}^{\infty} P\{y(j)=k\} = \frac{\pi(1)}{1-(1-\pi(1))} = 1$$

$$E[y(j)] = \sum_{k=1}^{\infty} kP\{y(j)=k\} = \frac{1}{1-\pi(0)} = \frac{1}{\pi(1)}$$

Now assume that the Markov structure of the zero's and ones is described by the transition matrix

$$\underline{P} = \begin{Bmatrix} P(0,0) & P(0,1) \\ P(1,0) & P(1,1) \end{Bmatrix} = \begin{Bmatrix} P+(1-\rho)\pi(1) & (1-\rho)\pi(0) \\ (1-\rho)\pi(1) & \rho+(1-\rho)\pi(0) \end{Bmatrix}$$

Here $P(m,n) = P\{x(i+1)=n \mid x(i)=m\}$, and we have parameterized the chain in terms of the stationary probability of a one or zero, and a correlation parameter $0\leq\rho<1$. Note that there are only two degrees of freedom in the stochastic

matrix, since rows must sum to 1, and there is only one degree of freedom if the stationary probability $\pi(0)=1-\pi(1)$ is fixed. Note that the stationary probabilities in the 2-state case are given by

$$\pi(0) = \frac{P(1,0)}{2-P(0,0)-P(1,1)} \qquad \pi(1) = \frac{P(0,1)}{2-P(0,0)-P(1,1)}$$

We now define the runs of ones or zeros i.e. for $\ell=0$ or $\ell=1$, let

$$T_\ell = \inf\{n\geq 1: \ x(i+n) \neq \ell\}-1 \ ,$$

the length of a run of ℓ's, starting after time i, where the length can be 0,1,2,... .

For example if $x(i+1)=1$, then the length of runs of zeros starting after time i is zero, the length of runs of ones is at least one long. Note that it is possible to talk of a conditional runs structure, i.e. the length of a run of ones which is given to start after time i. The run length is then at least one long.

Now the probability of a run T_ℓ having length greater than k is, using the Markov property,

$$P\{T_\ell\geq k\}= P\{x(i+1)=x(i+2)=\ldots x(i+k)=\ell\}=\pi(\ell)[P(\ell,\ell)]^{k-1} \qquad k=1,\ldots$$

and $$P\{T_\ell =0\} = 1-\pi(\ell) \ .$$

Thus, the run lengths are geometrically distributed and

$$E[T(\ell)] = \sum_{k=1}^{\infty} P\{T_\ell\geq k\} = \frac{\pi(\ell)}{1-P(\ell,\ell)} = \frac{\pi(\ell)}{(1-\rho)[1-\pi(\ell)]}$$

Note that $\rho=0$ gives the independence case, and while the runs of ones or zeros are geometrically distributed for both the independence or Markov dependent model, the mean run length is always longer for the Markov dependence, since

$$\frac{\pi(\ell)}{(1-\rho)[1-\pi(\ell)]} \geq \frac{\pi(\ell)}{[1-\pi(\ell)]} \qquad 0\leq\rho<1$$

Thus, we could use the distributional properties of the runs to (1) check that either hypothesis is tenable or (2) if so, compare the estimated run lengths with the mean length $\hat{\pi}(\ell)/[1-\hat{\pi}(\ell)]$ predicted by the independence assumption. If the run lengths are not geometric, than another model must be postulated.

Note that when this mean time-between-errors is large as it is for telephone data 1 (figure 1; $E[y(j)]=$ 1,548) the discreteness of the time scale can be ignored and the geometric distribution is indistinguishable from its continuous time analog, the exponential distribution.

That is approximation of the geometric distribution by an exponential distribution is valid can be seen from the fact that there are 672 errors ($x(i)$'s equal to one) in 1,106,148 transmitted bits, so that an estimate of $\pi(1)$, which is the maximum likelihood estimate under the independence hypothesis, is

$$\hat{\pi}(1) = \frac{\#\ x(i)\text{'s} = 1}{\text{total \# bits transmitted}} = \frac{\#\ x(i)\text{'s} = 1}{\#\ x(i)\text{'s}=1+\#x(i)\text{'s}=0}$$

In the present data

$$\hat{\pi}(1) = \frac{672}{1,106,148} = .0006075$$

Now this geometric hypothesis will be examined, but it is clear from figure 1 that the hypothesis is not true. The distribution is in fact highly skewed and has been examined by Lewis & Cox, 1966.

An alternative model to independent bit errors is that the dependence structure is Markovian. One could examine this hypothesis with time-series methods but a method which is adaptable for use with the histogram routine and which examines both the independence and Markov assumptions is to look at runs of ones and zeros in the x(i). Under both hypothesis these runs have geometrically distributed lengths.

The alternating conditional runs of ones for telephone data 1 are shown in figure 15 and for runs of zeros are shown in figure 16. Also, HISTLIST was used on the conditional runs and figure 17 shows the runs of ones and figure 18 shows the runs of zero.

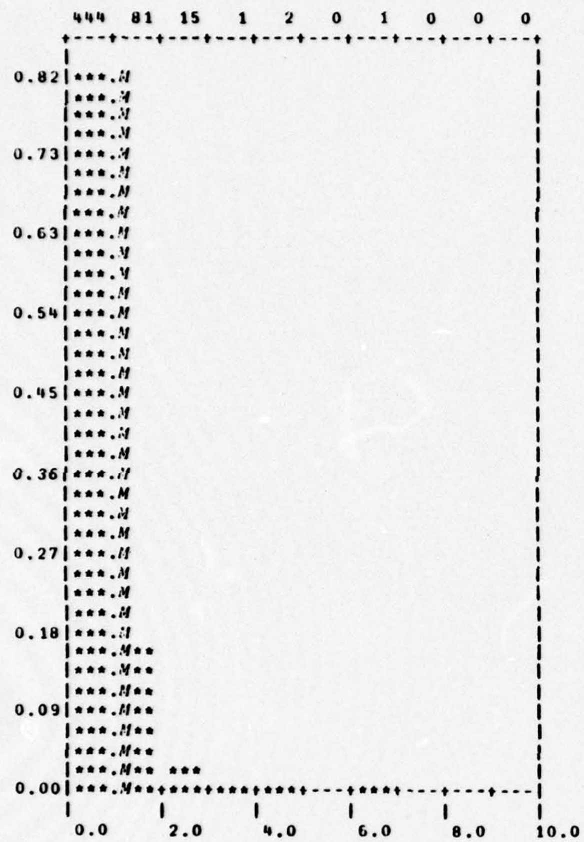
To test the hypothesis that the runs of ones in telephone data 1 is geometrically distributed the following was done. Using figures 15 and 17 the following data was obtained:

MEAN = 1.235294	# of runs = 1 = 444
VARIANCE = .346008	# of runs = 2 = 81
	# of runs = 3 = 15
	# of runs = 4 = 1
	# of runs = 5 = 2
	# of runs $\geq$ 6 = 1

FIGURE 15

RUNS OF ONE FOR TELEPHONE DATA 1

FREQUENCIES　　　SAMPLE SIZE = 544



CELL WIDTH = 1.000000E00

CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION	
MEAN	1.235294E00	VARIANCE	3.460080E-01	M3	7.990531E-01	MINIMUM	1.000000E00
MEDIAN	1.000000E00	STD DEV	5.882245E-01	M4	3.219155E00	.10 QUANTILE	1.000000E00
TRIMEAN	1.000000E00	COEF VAR	4.761817E-01	SKEWNESS	3.925965E00	.25 QUANTILE (HINGE)	1.000000E00
MIDMEAN	1.000000E00	MEAN DEV	2.352941E-01	KURTOSIS	2.388868E01	.50 QUANTILE (MEDIAN)	1.000000E00
MIDRANGE	4.000000E00	RANGE	6.000000E00	BETA1	7.946519E-01	.75 QUANTILE (HINGE)	1.000000E00
GEOM MEAN	1.156667E00	MIDSPREAD	0.000000E00	BETA2	3.196802E00	.90 QUANTILE	2.000000E00
HARM MEAN	1.109541E00					MAXIMUM	7.000000E00

FIGURE 16

RUNS OF ZERO FOR TELEPHONE DATA 1

FREQUENCIES SAMPLE SIZE = 544

CELL WIDTH = 7.142857E02

CENTRAL TENDENCY		SPREAD		HIGHER CENTRAL MOMENTS		DISTRIBUTION	
MEAN	1.911270E03	VARIANCE	5.906497E07	M3	2.911020E12	MINIMUM	1.000000E00
MEDIAN	2.400000E01	STD DEV	7.685374E03	M4	1.855145E17	.10 QUANTILE	1.000000E00
TRIMEAN	5.087500E01	COEF VAR	4.021082E00	SKEWNESS	6.412838E00	.25 QUANTILE (HINGE)	6.000000E00
MIDMEAN	3.894301E01	MEAN DEV	1.903406E03	KURTOSIS	5.017627E01	.50 QUANTILE (MEDIAN)	2.400000E01
MIDRANGE	4.299650E04	RANGE	8.599100E04	BETA1	2.894986E12	.75 QUANTILE (HINGE)	1.495000E02
GEOM MEAN	3.822483E01	MIDSPREAD	1.435000E02	BETA2	1.841888E17	.90 QUANTILE	1.546000E03
HARM MEAN	5.585516E00					MAXIMUM	8.599200E04

```
      HISTLIST
HISTLIST PRINTS THE SERIAL NUMBER OF THE  COMPRESSED
DATA, THE ORDERED DATA COMPRESSED, AND THE NUMBER OF
LIKE  OCCURENCES.  ENTER YOUR  DATA  IN VECTOR FORM.
[]:
      ONE

IF YOU WANT TO  TITLE  YOUR DATA  TYPE  YOUR  TITLE.
IF YOU DO NOT WANT A TITLE  JUST  HIT  THE  CARRIAGE
RETURN.

RUNS OF ONE

IF YOU WANT YOUR  OUTPUT TO APPEAR  ON  THE  OFFLINE
PRINTER TYPE  1  . IF YOU WANT YOUR OUTPUT TO APPEAR
ON YOUR TERMINAL TYPE  0  .
[]:
      0


RUNS OF ONE

SERIAL NUMBER    ORDERED DATA     NUMBER OF OCCURENCES                                   PER CENT

          1          1.000000                 444  *************************************************   0.816
        445          2.000000                  81  *********                                           0.149
        526          3.000000                  15  **                                                  0.028
        541          4.000000                   1                                                      0.002
        542          5.000000                   2                                                      0.004
        544          7.000000                   1                                                      0.002
```

FIGURE 17

FIGURE 18A

RUNS OF ZERO

SERIAL NUMBER	ORDERED DATA	NUMBER OF OCCURENCES		PER CENT
1	1.000000	54	******	0.099
55	2.000000	28	***	0.051
83	3.000000	22	**	0.040
105	4.000000	17	**	0.031
122	5.000000	11	*	0.020
133	6.000000	10	*	0.018
143	7.000000	12	*	0.022
155	8.000000	14	**	0.026
169	9.000000	9	*	0.017
178	10.000000	10	*	0.018
188	11.000000	11	*	0.020
199	12.000000	6	*	0.011
205	13.000000	6	*	0.011
211	14.000000	6	*	0.011
217	15.000000	8	*	0.015
225	16.000000	8	*	0.015
233	17.000000	5	*	0.009
238	18.000000	12	*	0.022
250	19.000000	1		0.002
251	20.000000	5	*	0.009
256	21.000000	5	*	0.009
261	22.000000	3		0.006
264	23.000000	7	*	0.013
271	24.000000	3		0.006
274	25.000000	3		0.006
277	26.000000	2		0.004
279	27.000000	3		0.006
282	28.000000	5	*	0.009
287	29.000000	6	*	0.011
293	30.000000	4		0.007
297	31.000000	4		0.007
301	32.000000	2		0.004
303	33.000000	4		0.007
307	34.000000	3		0.006
310	36.000000	2		0.004
312	37.000000	2		0.004
314	38.000000	1		0.002
315	39.000000	2		0.004
317	40.000000	2		0.004
319	42.000000	1		0.002
320	43.000000	1		0.002
321	44.000000	4		0.007
325	45.000000	3		0.006
328	46.000000	1		0.002
329	47.000000	2		0.004
331	48.000000	1		0.002
332	49.000000	1		0.002
333	50.000000	2		0.004
335	51.000000	1		0.002
336	52.000000	1		0.002
337	53.000000	3		0.006
340	54.000000	2		0.004
342	[illegible].000000	2		0.004
344	57.000000	2		0.004
346	58.000000	1		0.002
347	61.000000	2		0.004
349	62.000000	3		0.006
352	63.000000	1		0.002
353	64.000000	2		0.004
355	65.000000	1		0.002
356	67.000000	3		0.006
359	68.000000	1		0.002
360	69.000000	1		0.002
361	72.000000	1		0.002
362	73.000000	2		0.004
364	74.000000	1		0.002
365	78.000000	3		0.006
368	82.000000	1		0.002
369	83.000000	1		0.002
370	85.000000	1		0.002
371	87.000000	2		0.004
373	88.000000	1		0.002
374	89.000000	2		0.004
376	90.000000	1		0.002
377	92.000000	1		0.002
378	94.000000	1		0.002
379	97.000000	1		0.002
380	98.000000	1		0.002
381	105.000000	1		0.002
382	108.000000	1		0.002
383	110.000000	1		0.002
384	111.000000	2		0.004
386	112.000000	1		0.002
387	115.000000	1		0.002

FIGURE 18B

388	116.000000	1	0.002
389	118.000000	1	0.002
390	119.000000	4	0.007
394	120.000000	2	0.004
396	121.000000	2	0.004
398	122.000000	4	0.007
402	123.000000	3	0.006
405	127.000000	1	0.002
406	134.000000	1	0.002
407	141.000000	1	0.002
408	147.000000	1	0.002
409	152.000000	1	0.002
410	155.000000	1	0.002
411	157.000000	1	0.002
412	160.000000	1	0.002
413	164.000000	1	0.002
414	174.000000	1	0.002
415	175.000000	1	0.002
416	176.000000	1	0.002
417	182.000000	1	0.002
418	185.000000	1	0.002
419	186.000000	1	0.002
420	191.000000	1	0.002
421	192.000000	1	0.002
422	201.000000	1	0.002
423	216.000000	2	0.004
425	223.000000	1	0.002
426	225.000000	1	0.002
427	227.000000	3	0.006
430	230.000000	1	0.002
431	233.000000	1	0.002
432	236.000000	1	0.002
433	238.000000	2	0.004
435	239.000000	4	0.007
439	240.000000	1	0.002
440	243.000000	1	0.002
441	247.000000	1	0.002
442	248.000000	1	0.002
443	250.000000	1	0.002
444	251.000000	1	0.002
445	259.000000	1	0.002
446	278.000000	1	0.002
447	285.000000	1	0.002
448	296.000000	1	0.002
449	302.000000	1	0.002
450	311.000000	1	0.002
451	317.000000	1	0.002
452	346.000000	1	0.002
453	[illegible]	1	0.002
454	359.000000	1	0.002
455	363.000000	1	0.002
456	365.000000	1	0.002
457	389.000000	1	0.002
458	393.000000	1	0.002
459	434.000000	1	0.002
460	459.000000	1	0.002
461	464.000000	1	0.002
462	472.000000	1	0.002
463	479.000000	1	0.002
464	490.000000	1	0.002
465	548.000000	1	0.002
466	599.000000	2	0.004
468	608.000000	1	0.002
469	620.000000	1	0.002
470	710.000000	1	0.002
471	816.000000	1	0.002
472	835.000000	1	0.002
473	926.000000	1	0.002
474	1123.000000	1	0.002
475	1149.000000	1	0.002
476	1269.000000	1	0.002
477	1288.000000	1	0.002
478	1297.000000	1	0.002
479	1304.000000	1	0.002
480	1327.000000	1	0.002
481	1347.000000	1	0.002
482	1354.000000	1	0.002
483	1411.000000	1	0.002
484	1428.000000	1	0.002
485	1468.000000	1	0.002
486	1492.000000	1	0.002
487	1509.000000	1	0.002
488	1518.000000	1	0.002
489	1546.000000	1	0.002
490	1632.000000	1	0.002
491	1766.000000	1	0.002
492	2071.000000	1	0.002
493	2482.000000	1	0.002
494	2805.000000	1	0.002
495	2961.000000	1	0.002
496	3025.000000	1	0.002

FIGURE 18C

497	3200.000000	1	0.002
498	3592.000000	1	0.002
499	3684.000000	1	0.002
500	3951.000000	1	0.002
501	4156.000000	1	0.002
502	4988.000000	1	0.002
503	6207.000000	1	0.002
504	7613.000000	1	0.002
505	8321.000000	1	0.002
506	9014.000000	1	0.002
507	9624.000000	1	0.002
508	9805.000000	1	0.002
509	9817.000000	1	0.002
510	10193.000000	1	0.002
511	10397.000000	1	0.002
512	10490.000000	1	0.002
513	10938.000000	1	0.002
514	11279.000000	1	0.002
515	13446.000000	1	0.002
516	14384.000000	1	0.002
517	15194.000000	1	0.002
518	15293.000000	1	0.002
519	15503.000000	1	0.002
520	15846.000000	1	0.002
521	15867.000000	1	0.002
522	16279.000000	1	0.002
523	16298.000000	1	0.002
524	16360.000000	1	0.002
525	16407.000000	1	0.002
526	16816.000000	1	0.002
527	17173.000000	1	0.002
528	17666.000000	1	0.002
529	18217.000000	1	0.002
530	18648.000000	1	0.002
531	19460.000000	1	0.002
532	21847.000000	1	0.002
533	23497.000000	1	0.002
534	24891.000000	1	0.002
535	26442.000000	1	0.002
536	30973.000000	1	0.002
537	35643.000000	1	0.002
538	38002.000000	1	0.002
539	40130.000000	1	0.002
540	47119.000000	1	0.002
541	47591.000000	1	0.002
542	61709.000000	1	0.002
543	69174.000000	1	0.002
544	85992.000000	1	0.002

If the runs of ones are geometric then prob{x(i)=k} = $(1-p)p^{k-1}$ k=1,2,... . Thus, this is the "geometric plus one" distribution.

$$\mu = E[X] = \frac{1}{(1 - p)}$$

$$\sigma^2 = VAR[X] = \frac{1}{(1 - p)^2}$$

$$C(X) = \frac{VAR[X]^{\frac{1}{2}}}{E[X]} = p^{\frac{1}{2}}$$

To find p set E[X] = 1.235294 = 1/(1-p)

$$p = \underline{.1904761}$$

Therefore, if the data is "geometric plus one" then

$$\text{EXPECTED VAR}[X] = .1904761/(.8095329)^2$$
$$= \underline{.2906572}$$

Thus, the expected variance is .2906572 and the observed variance from HIST is .3460080 . Also, the expected coefficient of variance is

$$\text{EXPECTED } C(X) = (.1904761)^{\frac{1}{2}} = \underline{.4364356}$$

And, the observed coefficient of variation is .4761817 .

Therefore, at this point there seems to be a fairly close agreement between the runs of one and a "geometric plus one" distribution with p = .1904761 .

As further proof a Chi-square test for goodness of fit was run on the runs. By using the formula

$$\text{prob } \{X = x\} = (1-p)p^{x-1} \quad \text{for } x=1,2,3,4,5,...$$

PROBABILITY	EXPECTED		OBSERVED	
$P(X=1) = .8095239$	440.38		444	
$P(X=2) = .1541949$	83.88		81	
$P(X=3) = .0293704$	15.98	19.74	15	19
$P(X=4) = .0055943$	3.04		1	
$P(X=5) = .0010655$	.58		2	
$P(X\geq 6) = .0002510$	.14		1	

Note, to use Chi-square not more than 20% of the cells should have expected frequencies less than 5 and no cell should have an expected frequency less than one. Therefore, the above frequencies must be combined into 3 cells.

$$\chi^2 = \sum_{i=1}^{3} \frac{(obs_i - ex_i)^2}{ex_i} = \underline{.1562799}$$

And, $\chi^2_{.05,2} = 5.99$. Thus, the null hypothesis that the runs of one are "geometric plus one" with p = .1904761 can not be rejected.

A similar procedure was done with the runs of greater than one. By using figure 15 the following information can be obtained:

MEAN = 1911.27
VARIANCE = 59,064,970
COEF.VAR.= 4.021082

And, by using the same method as previously done and solving for *p* *one gets* $p = \underline{.9994767}$.

$$\text{EXPECTED VAR}[X] = .9994767/(.0005233)^2 = \underline{3,651,213}$$

This expected variance differs greatly from the observed variance. Also, the expected coefficient of variation is

computed to be

$$\text{EXPECTED } C(X) = (.9994767)^{\frac{1}{2}} = \underline{.9997383}$$

This compares with the observed coefficient of variation of 4.021082 . Because of the gross departures of the variance and the coefficient of variation in the geometric hypothesis, one can conclude that the runs of length greater than 1 are not geometrically distributed.

IX. DOCUMENTATION ON ROUTINES

A. LOCATION IN APL LIBRARY

The descriptions and routines that have been presented are all available in the APL workspace library 2 DATALFNS . Providing the user is properly logged on the terminal and in the APL mode, all that is necessary is to type)LOAD 2 DATALFNS . If the user then types DESCRIBE, a short description of the six routines presented and instructions on how to obtain the detailed information that is available in each of the "HOW" variables would be printed.

B. WORKSPACE LOADING PROCEDURES

Each of the routines was designed to stand alone. That is, if the user desires just to use HIST , all that is necessary is to type)COPY 2 DATALFNS HISTGRP into a clear workspace. HISTGRP contains the principal routine HIST and only the additional routines necessary for HIST to operate. Thus, the user does not clutter his workspace with any unneeded functions. It is this group structure that maintains the orderliness of the workspace. And, the ability to copy a particular group into a clear workspace provides more space for data and executions of the functions.

The following is the group structure in library 2 DATALFNS .

GROUP	PRINCIPAL ROUTINE	OTHER NECESSARY ROUTINES	VARIABLES
HISTGRP	HIST	APLNAME,APLOT,AUTOS, CMS,DFT,ECDF,ECODE, EFT,OF,OUT,WRITE	
HISTLISTGRP	HISTLIST	APLNAME,CMS,ECODE, DFT,OF,OUT,WRITE	
HISTSGRP	HISTS	DFT,EFT	
HISTJACKGRP	HISTJACK	DFT,EFT,TOT	
EXPONPGRP	EXPONP	AND,AUTOSCALE, INITIAL,MPLOT,MSGS, VS,MULTIPLOT,SET∆AP, TICMARK	BS
NORMPGRP	NORMP	AND,AUTOSCALE, INITIAL,MPLOT,MSGS, VS,MULTIPLOT,SET∆AP, TICMARK	BS
DESCGRP (Descriptive group)			DESCRIBE,HISTHOW HISTHOW,HISTLIST-HOW,HISTJACKHOW, EXPONPHOW,NORMPHOW
VARIGRP (Variable group)			TELDAT1,TELDAT2, YROVR

C. ROUTINE LISTING

The above mentioned routines were either created by the author, adapted from existing fortran routine HISTG/F , or borrowed from the current APL library to supplement the author created routines.

1. Author Created Routines

HISTLIST, HISTS, HISTJACK, EXPONP, NORMP, APLOT, AUTOS, OUT, TOT

2. Adapted from Fortran Library Routine HISTG/F

HIST, ECDF

3. Borrowed Routines to Supplement Author Created Routines

AND, APLNAME, AUTOSCALE, CMS, DFT, ECODE, EFT, INITIAL, MPLOT, MSGS, MULTIPLOT, NDTRI, OF, SET∆AP, TICMARK, VS, WRITE

X. COMPUTER LISTING OF ALL ROUTINES

```
      ∇HIST[⎕]∇
    ∇ HIST;N;V;X;T;A;APLN;B;C;D;DELTA;F;HSCALE;M1;M2;N;OL;SUM;SUMA;SUMB;TT;XLABEL;Z;Δ1;Δ2;Δ3;Δ4;Δ41;Δ42;B1;B2;B3;C;B4;
      T1
[1]   'ENTER DATA IN VECTOR FORM'
[2]   X←⎕
[3]   ' '
[4]   'IF YOU ALSO WANT A SMOOTHED EMPIRICAL DENSITY FUNCTION ENTER'
[5]   'A   1  .  IF YOU DO NOT WANT IT ENTER A   0  .'
[6]   B←⎕
[7]   ' '
[8]   'IF  YOU WANT  TO  TITLE  YOUR  HISTOGRAM  TYPE  YOUR  TITLE.'
[9]   'IF  YOU  DO  NOT WANT A TITLE JUST HIT THE CARRIAGE  RETURN.'
[10]  ' '
[11]  TT←⍞
[12]  ' '
[13]  'IF YOU WANT TO SET THE NUMBER OF CELLS AND THE  SCALE  ENTER'
[14]  'FIRST THE  NUMBER OF CELLS (AN INTEGER BETWEEN  10  AND  28)'
[15]  'FOLLOWED BY A SPACE AND THEN  YOUR LEFT SCALE POINT FOLLOWED'
[16]  'BY A SPACE AND THEN YOUR RIGHT SCALE POINT.  HOWEVER, IF YOU'
[17]  'WANT HIST TO AUTOMATICALLY SCALE ENTER   0 0 0  .'
[18]  A←⎕
[19]  ' '
[20]  'GIVEN THAT YOU HAVE SET YOUR  OWN  SCALE,  TO  INCLUDE  DATA'
[21]  'POINTS THAT MIGHT  BE OUTSIDE YOUR SCALE LIMITS  IN  THE END'
[22]  'CELLS, TYPE   1  .  IF YOU DESIGNATED  AUTOSCALE  ALSO, TYPE'
[23]  '1  .  IF HOWEVER, YOU DO NOT WANT THE DATA OUTSIDE THE SCALE'
[24]  'LIMITS INCLUDED IN THE HISTOGRAM, TYPE   0  .'
[25]  Z←⎕
[26]  ' '
[27]  'IF YOU WANT YOUR OUTPUT TO APPEAR ON  THE  OFFLINE  PRINTER.'
[28]  'TYPE   1  .  IF YOU  WANT YOUR  OUTPUT  TO  APPEAR  ON  YOUR'
[29]  'TERMINAL, TYPE   0  .  (NOTE IF YOU TYPED   0  BE SURE YOUR'
[30]  'TERMINALS CARRIAGE PAGE SETTING IS  ON  THE  MAXIMUM  WIDTH)'
[31]  OL←⎕
[32]  D← 37 ¯3 4
[33]  C←29⍴0
[34]  →(OL=0)/TABJ
[35]  APLN←APLNAME 'HIST APLPF P1 V'
[36]  CMS 'ERASE HIST APLPF'
[37]  (20↑'1') WRITE APLN
[38]  D← 55 ¯5 6
[39] TABJ:→(Z≠0)/TAB2A
[40]  X←((X≥A[2])∧(X≤A[3]))/X
[41] TAB2A:C[12]←(C[27]←X[⍴X])-C[21]←(X←X[⍋X])[1]
[42]  →(A[1]≠0∧A[2]≠0∧A[3]≠0)/TAB1A
[43]  AUTOS
```

```
[44] TAB1A:→(A[1]<10)/LAST
[45]  →(A[1]>28)/LAST
[46]  DELTA←(HSCALE←A[3]-A[2])÷A[1]
[47]  XLABEL←A[2],(A[2]+DELTA×⍳A[1])
[48]  F←+/(⍳A[1])∘.=⌈(X-A[2])÷DELTA
[49]  F[1]←(+/(X≤XLABEL[1]))+F[1]
[50]  F[A[1]]←(+/(X>XLABEL[A[1]+1]))+F[A[1]]
[51]  C[9]←(C[8]←(+/(X-C[1]←(+/X)÷N)*2)÷(N←ρX)-1)*0.5
[52]  C[2]←0.5×+/X[(⌈N÷2),1+⌊N÷2]
[53]  C[11]←(+/|X-C[2])÷N
[54]  C[22]←X[⌊(N÷10)]
[55]  C[29]←N-C[28]←(⌊N÷4)
[56]  M2←1+M1←(1-((4|N)÷2))
[57]  C[23]←(X[C[28]+1]+(X[C[28]]×M1))÷M2
[58]  C[24]←C[2]
[59]  C[25]←(X[C[29]]+(X[C[29]+1]×M1))÷M2
[60]  C[26]←X[⌊(0.9×N)]
[61]  C[13]←C[25]-C[23]
[62]  C[5]←(C[27]+C[21])×0.5
[63]  C[3]←0.25×(C[23]+C[24]+C[24]+C[25])
[64]  C[15]←((+/((X-C[1])*3))×N)÷((N-1)×(N-2))
[65]  C[19]←(+/((X-C[1])*3))÷N
[66]  C[16]←(+/((X-C[1])*4))×(3+N×(N-2))÷((N-1)×(N-2)×(N-3))
[67]  C[20]←(+/((X-C[1])*4))÷N
[68]  C[16]←C[16]-(C[8]×C[8]×3×(N-1)×(N+N-3))÷(N×(N-2)×(N-3))
[69]  C[18]←¯3+C[16]÷(C[8]×C[8])
[70]  C[29]←N+1-C[28]←2+⌊N÷4
[71]  SUM←C[23]+C[25]
[72]  SUMA←+/X[⍳(C[28]-1)]
[73]  SUMB←+/X[⍳(C[29])]
[74]  SUM←SUM+(SUMB-SUMA)
[75]  C[4]←SUM÷(3+C[29]-C[28])
[76]  C[17]←C[15]÷C[9]*3
[77]  C[6]←C[7]←0
[78]  SUMA←5
[79]  SUMB←7
[80]  →(X[1]≤0)/TAB
[81]  C[6]←*((+/(⍟X))÷N)
[82]  C[7]←N÷(+/(÷X))
[83]  SUMA←SUMB←7
[84] TAB:→(N>300)/TAB4
[85]  →(N>C[14]←ρΔ1←((ρV)ρ(⍳M)≤1)/V←X[(V=M←⌈/V←+/X∘.=X)/⍳ρX])/TAB2
[86] TAB4:C[14]←0
[87]  SUMB←6
[88] TAB2:C[10]←0
```

```
→(|C[1]<1E 30)/TAB3
 C[10]←C[9]÷|C[1]
TAB3:T1←'CENTRAL TENDENCY                    SPREAD                          HIGHER CENTRAL MOMENTS      DISTRIBUTION'
 A1←'MEAN       MEDIAN      TRIMEAN    MIDMEAN    MIDRANGE   GEOM MEAN  HARM MEAN  '
 A2←'VARIANCE   STD DEV     COEF VAR   MEAN DEV   RANGE      MIDSPREAD  MODE       '
 A3←'M3         M4          SKEWNESS  KURTOSIS  BETA1       BETA2    '
 A41←'MINIMUM                .10 QUANTILE           .25 QUANTILE (HINGE)   .50 QUANTILE (MEDIAN)  '
 A42←'.75 QUANTILE (HINGE)   .90 QUANTILE           MAXIMUM                 '
 A4←A41,A42
 B1← 13 7 EFT B1←((SUMA),1)ρB1←C[1],C[2],C[3],C[4],C[5],C[6],C[7]
 B2← 13 7 EFT B2←((SUMB),1)ρB2←C[8],C[9],C[10],C[11],C[12],C[13],C[14]
 B3←,B3← 13 7 EFT B3← 6 1 ρB3←C[15],C[16],C[17],C[18],C[19],C[20]
 B4← 13 7 EFT B4← 7 1 ρB4←C[21],C[22],C[23],C[24],C[25],C[26],C[27]
 A1←((SUMA),11)ρA1
 A2←((SUMB),11)ρA2
 A3← 7 10 ρA3
 A4← 7 23 ρA4
 C← 7 4 ρC←' '
 APLOT
TAB5:OL OUT T←'                    '
 OL OUT 'CELL WIDTH = ', 13 7 EFT DELTA
 B3← 7 13 ρB3←B3,T
 OL OUT 2 7 ρT
 OL OUT T1
 OL OUT T
 →(SUMA=5)/TAB6
 →(SUMB=6)/TAB7
TAB8:OL OUT A1,B1,C,A2,B2,C,A3,B3,C,A4,B4
 →(OL=0)/0
 CMS 'FINIS HIST APLPF'
 CMS 'O PRINTCC HIST APLPF'
 →(0≠ECODE)/EX1
 'HISTOGRAM SENT TO PRINTER'
 CMS 'ERASE HIST APLPF'
 →0
EX1:0,ρ⍞←'PRINTING FAILED.  TRY AGAIN OR SEE APL PROGRAMMER.'
TAB6:A1← 7 11 ρA1
 B1← 7 13 ρB1
 A1[6;]←A1[7;]←B1[6;]←B1[7;]←' '
 →(SUMB=7)/TAB8
TAB7:A2← 7 11 ρA2
 B2← 7 13 ρB2
 A2[7;]←B2[7;]←' '
 →TAB8
LAST:'NUMBER OF CELLS GIVEN IS NOT PERMISSABLE'
∇
```

```
      ∇HISTLIST[⎕]∇
    ∇ HISTLIST;X;A;B;C;D;E;F;G;I;J;K;N;O;S;TT;APLN;OL
[1]    'HISTLIST PRINTS THE SERIAL NUMBER OF THE  COMPRESSED'
[2]    'DATA. THE ORDERED DATA COMPRESSED, AND THE NUMBER OF'
[3]    'LIKE  OCCURENCES.  ENTER YOUR  DATA  IN VECTOR FORM.'
[4]    X←⎕
[5]    ' '
[6]    'IF YOU WANT TO  TITLE  YOUR DATA  TYPE  YOUR  TITLE.'
[7]    'IF YOU DO NOT WANT A TITLE  JUST  HIT  THE  CARRIAGE'
[8]    'RETURN.'
[9]    ' '
[10]   TT←⍞
[11]   ' '
[12]   'IF YOU WANT YOUR  OUTPUT TO APPEAR  ON  THE  OFFLINE'
[13]   'PRINTER TYPE  1  . IF YOU WANT YOUR OUTPUT TO APPEAR'
[14]   'ON YOUR TERMINAL TYPE  0  .'
[15]   OL←⎕
[16]   →(OL=0)/TABA
[17]   APLN←APLNAME 'HISTLIST APLPF P1 V'
[18]   CMS 'ERASE HISTLIST APLPF'
[19]   (20↑'1') WRITE APLN
[20]  TABA:→(TT=0)/TAB10
[21]   OL OUT 2 7 ρ' '
[22]   OL OUT TT
[23]   OL OUT 2 7 ρ' '
[24]  TAB10:X←X[⍋X]
[25]   C←(ρX)ρ1
[26]   K←1
[27]   E←1ρ1
[28]   J←0
[29]   F←⍳0
[30]  TAB2:J←J+1
[31]   →(X[J]=X[J+1])/TAB1
[32]   K←K+1
[33]   E←E,(J+1)
[34]   F←F,X[J]
[35]   →((J+1)=ρX)/TAB3
[36]   →TAB2
[37]  TAB1:C[K]←C[K]+1
[38]   →((J+1)=ρX)/TAB3
[39]   →TAB2
```

```
[40] TAB3:F←F,X[J+1]
[41]  A←'SERIAL NUMBER    ORDERED DATA      NUMBER OF OCCURENCES'
[42]  B←⌊(0.5+60×((⌈/C)÷(ρX)))
[43]  D←8⌈(B+2)
[44]  B←(2⌈(B-4))ρ' '
[45]  OL OUT A,B,'PER CENT'
[46]  OL OUT '               '
[47]  J←0
[48] TAB4:J←J+1
[49]  G←⌊(0.5+60×I←(C[J]÷(ρX)))
[50]  B←G↓Dρ' '
[51]  G←Gρ'*'
[52]  I←3 DFT I
[53]  S← 10 0 DFT E[J]
[54]  O← 16 6 DFT F[J]
[55]  N← 10 0 DFT C[J]
[56]  OL OUT S,'      ',O,'      ',N,'  ',G,B,I
[57]  →(J=K)/TAB5
[58]  →TAB4
[59] TAB5:→(OL=0)/0
[60]  CMS 'FINIS HISTLIST APLPF'
[61]  CMS 'O PRINTCC HISTLIST APLPF'
[62]  →(0≠ECODE)/EX1
[63]  'HISTLIST SENT TO PRINTER'
[64]  CMS 'ERASE HISTLIST APLPF'
[65]  →0
[66] EX1:0,ρ⎕←'PRINTING FAILED.  TRY AGAIN OR SEE APL PROGRAMMER.'
    ∇
```

```
      ∇HISTS[⎕]∇
    ∇ HISTS;X;P;SE;ARRAY;J;FN;A;B;C;D;E;F;G;H;I;K;SD;VAR;MED;MIN;MAX;SDS;STS;KURT;SKEW;CVAR;MEAN;VRS;MNS;SZ;M3;M4;N
[1]   'TYPE THE NUMBER OF SECTIONS YOU  DESIRE   ( INTEGER'
[2]   'BETWEEN 2 AND 28 )  BE SURE TO PICK YOUR NUMBER OF'
[3]   'SECTIONS SO AS TO  MINIMIZE  THE  NUMBER  OF  DATA'
[4]   'POINTS THAT WILL  HAVE  TO  BE  DISCARDED.  (HISTS'
[5]   'PLACES THE DATA INTO THE EQUAL NUMBER OF  SECTIONS'
[6]   'YOU  INDICATE  DISCARDING  ANY  DATA  LEFT  OVER )'
[7]   SE←⎕
[8]   ' '
[9]   'ENTER  YOUR  DATA  TO BE SECTIONED IN VECTOR  FORM'
[10]  X←⎕
[11]  ' '
[12]  P←0
[13]  SDS←VRS←MNS←STS←7⍴0
[14] TAB10:SZ←⌊(⍴X)÷SE
[15]  MEAN←VAR←SD←CVAR←SKEW←MED←MIN←KURT←MAX←(SE)⍴0
[16]  ARRAY←((SE),(SZ))⍴X
[17]  J←0
[18] TAB3:J←J+1
[19]  →(J>SE)/TAB2
[20]  MAX[J]←⌈/(ARRAY[J;])
[21]  MIN[J]←⌊/(ARRAY[J;])
[22]  SD[J]←(VAR[J]←(+/(ARRAY[J;]-MEAN[J]←(+/ARRAY[J;])÷N)*2)÷(N←SZ)-1)*0.5
[23]  FN←ARRAY[J;]
[24]  FN←FN[⍋FN]
[25]  MED[J]←0.5×(+/FN[(⌈N÷2),1+⌊N÷2])
[26]  M3←M4←(SE)⍴0
[27]  M3[J]←((+/((ARRAY[J;]-MEAN[J])*3))×N)÷((N-1)×(N-2))
[28]  M4[J]←(+/((ARRAY[J;]-MEAN[J])*4))×(3+N×(N-2))÷((N-1)×(N-2)×(N-3))
[29]  M4[J]←M4[J]-(VAR[J]×VAR[J]×3×(N-1)×(N+N-3))÷(N×(N-2)×(N-3))
[30]  SKEW[J]←M3[J]÷SD[J]*3
[31]  KURT[J]←¯3+M4[J]÷(VAR[J]×VAR[J])
[32]  CVAR[J]←SD[J]÷|MEAN[J]
[33]  →TAB3
[34] TAB2:→(P=1)/TAB12
[35]  ARRAY←MEAN,MED,VAR,SD,CVAR,SKEW,KURT
[36]  ARRAY←(7,(SE))⍴ARRAY
[37]  J←0
[38] TAB4:J←J+1
[39]  SDS[J]←(VRS[J]←(+/(ARRAY[J;]-MNS[J]←(+/ARRAY[J;])÷N)*2)÷(N←SE)-1)*0.5
[40]  STS[J]←SDS[J]÷((N)*0.5)
[41]  →(J=7)/TAB5
[42]  →TAB4
[43] TAB5:A←'SECTION     MEAN        MEDIAN      VARIANCE    STD DEV     COEF VAR    '
[44]  B←'SKEWNESS    KURTOSIS    MINIMUM     MAXIMUM     '
[45]  A,B
```

```
[46]  ' '
[47] TAB12:J←0
[48] TAB6:J←J+1
[49]  K← 2 0 DFT J
[50]  A← 11 5 EFT MEAN[J]
[51]  B← 11 5 EFT MED[J]
[52]  C← 11 5 EFT VAR[J]
[53]  D← 11 5 EFT SD[J]
[54]  E← 11 5 EFT CVAR[J]
[55]  F← 11 5 EFT SKEW[J]
[56]  G← 11 5 EFT KURT[J]
[57]  H← 11 5 EFT MIN[J]
[58]  I← 11 5 EFT MAX[J]
[59]  →(P=1)/TAB7
[60]  '   ',K,'       ',A,' ',B,' ',C,' ',D,' ',E,' ',F,' ',G,' ',H,' ',I
[61]  →(J=SE)/TAB11
[62]  →TAB6
[63] TAB11:P←SE+1
[64]  →TAB10
[65] TAB7: 2 7 ρ' '
[66]  'UNSECTIONED',' ',A,' ',B,' ',C,' ',D,' ',E,' ',F,' ',G,' ',H,' ',I
[67]  2 1 ρ' '
[68]  'SUMMARY FOR SECTIONED DATA'
[69]  ' '
[70]  '           MEAN        VARIANCE    STD DEV    STD÷(SECS)*.5'
[71]  ' '
[72]  J←0
[73] TAB8:J←J+1
[74]  A← 11 5 EFT MNS[J]
[75]  B← 11 5 EFT VRS[J]
[76]  C← 11 5 EFT SDS[J]
[77]  D← 11 5 EFT STS[J]
[78]  E←'MEAN         MEDIAN       VARIANCE    STD DEV     COEF VAR    SKEWNESS    KURTOSIS
[79]  E← 7 12 ρE
[80]  E[J;],A,' ',B,' ',C,' ',D
[81]  →(J=7)/0
[82]  →TAB8
    ∇
```

```
      ∇HISTJACK[⎕]∇
    ∇ HISTJACK;X;SEC1;PSV;SZ;A;B;C;J;G;ARRAY;BRRAY;K;FN;S;CVAR;KURT;MAX;MEAN;MEANS;MED;MIN;M3;M4;N;SD;SKEW;VAR;VARSA
[1]    'TYPE THE NUMBER  OF  GROUPS  YOU  DESIRE  (INTEGER
       BETWEEN  2  AND  50 )  BE SURE TO PICK YOUR NUMBER
       OF GROUPS SO AS TO MINIMIZE  THE  NUMBER  OF  DATA
       POINTS THAT WILL HAVE TO BE DISCARDED.   (HISTJACK
       PLACES THE DATA INTO THE EQUAL  NUMBER  OF  GROUPS
       YOU INDICATE DISCARDING ANY DATA LEFT OVER)'
[2]    SEC1←⎕
[3]    ' '
[4]    'ENTER YOUR DATA  TO  BE  JACKNIFED  IN VECTOR FORM'
[5]    X←⎕
[6]    ' '
[7]    MEANS←VARSA←S← 7 1 ⍴0
[8]    PSV←((7),(SEC1))⍴0
[9]    SZ←(⍴X)-(⌊((⍴X)÷SEC1))
[10]   MEAN←VAR←SD←CVAR←SKEW←MED←MIN←KURT←MAX←(SEC1+1)⍴0
[11]   ARRAY←(1,(⍴X))⍴X
[12]   J←G←1
[13]   B←⍴X
[14]  TAB3:→(J>(SEC1+1))/TAB2
[15]   MAX[J]←⌈/(ARRAY[G;])
[16]   MIN[J]←⌊/(ARRAY[G;])
[17]   SD[J]←(VAR[J]←(+/(ARRAY[G;]-MEAN[J]←(+/ARRAY[G;])÷N)*2)÷(N←B)-1)*0.5
[18]   FN←ARRAY[G;]
[19]   FN←FN[⍋FN]
[20]   MED[J]←0.5×(+/FN[(⌈N÷2),1+⌊N÷2])
[21]   M3←M4←(SEC1+1)⍴0
[22]   M3[J]←((+/((ARRAY[G;]-MEAN[J])*3))×N)÷((N-1)×(N-2))
[23]   M4[J]←(+/((ARRAY[G;]-MEAN[J])*4))×(3+N×(N-2))÷((N-1)×(N-2)×(N-3))
[24]   M4[J]←M4[J]-(VAR[J]×VAR[J]×3×(N-1)×(N+N-3))÷(N×(N-2)×(N-3))
[25]   SKEW[J]←M3[J]÷SD[J]*3
[26]   KURT[J]←¯3+M4[J]÷(VAR[J]×VAR[J])
[27]   CVAR[J]←SD[J]÷|MEAN[J]
[28]   G←(J←J+1)-1
[29]   B←SZ
[30]   →(G≥2)/TAB3
[31]   C←⌊(⍴X)÷A←SEC1
[32]   BRRAY←((A),(C))⍴X
[33]   ARRAY←((A),(SZ))⍴0
```

```
 TOT
 →TAB3
TAB2:PSV[1;]←(A×MEAN[1])-((A-1)×MEAN[1+⍳SEC1])
 PSV[2;]←(A×MED[1])-((A-1)×MED[1+⍳SEC1])
 PSV[3;]←(A×VAR[1])-((A-1)×VAR[1+⍳SEC1])
 PSV[4;]←(A×SD[1])-((A-1)×SD[1+⍳SEC1])
 PSV[5;]←(A×CVAR[1])-((A-1)×CVAR[1+⍳SEC1])
 PSV[6;]←(A×SKEW[1])-((A-1)×SKEW[1+⍳SEC1])
 PSV[7;]←(A×KURT[1])-((A-1)×KURT[1+⍳SEC1])
 MEANS←((+/PSV)÷A)
 VARSA←((+/PSV*2)-(((+/PSV)*2)÷SEC1))÷(SEC1-1)
 S←(VARSA÷SEC1)*0.5
 A←'GROUP         MEAN          MEDIAN        VARIANCE    STD DEV      COEF VAR    '
 C←'SKEWNESS      KURTOSIS      MINIMUM       MAXIMUM'
 ' '
 A,C
 ' '
 J←1
 A← 9 11 ρ' '
TAB4:J←J+1
 A[1;]← 11 5 EFT MEAN[J]
 A[2;]← 11 5 EFT MED[J]
 A[3;]← 11 5 EFT VAR[J]
 A[4;]← 11 5 EFT SD[J]
 A[5;]← 11 5 EFT CVAR[J]
 A[6;]← 11 5 EFT SKEW[J]
 A[7;]← 11 5 EFT KURT[J]
 A[8;]← 11 5 EFT MIN[J]
 A[9;]← 11 5 EFT MAX[J]
 K← 2 0 DFT(J-1)
 →(J=1)/TAB6
 '   ',K,'        ',A[1;],' ',A[2;],' ',A[3;],' ',A[4;],' ',A[5;],' ',A[6;],' ',A[7;],' ',A[8;],' ',A[9;]
 →(J=(SEC1+1))/TAB5
 →TAB4
TAB5:J←0
 →TAB4
```

```
[70] TAB6: 2 1 ρ' '
[71]  'UNGROUPED    ',A[1;],' ',A[2;],' ',A[3;],' ',A[4;],' ',A[5;],' ',A[6;],' ',A[7;],' ',A[8;],' ',A[9;]
[72]  2 1 ρ' '
[73]  'SUMMARY FOR JACKNIFED DATA'
[74]  ' '
[75]  '           JACKNIFE ESTIMATE    VARIANCE       (VAR÷GROUPS)*.5'
[76]  A←48ρ' '
[77]  A,'JACKNIFE ESTIMATE OF STD DEV'
[78]  A,'OF  MEAN  OF   PSEUDO-VALUES'
[79]  ' '
[80]  A←'MEAN      MEDIAN   VARIANCE STD DEV  COEF VAR SKEWNESS KURTOSIS '
[81]  A← 7 9 ρA
[82]  J←0
[83]  C← 3 11 ρ' '
[84] TAB7:J←J+1
[85]  C[1;]← 11 5 EFT MEANS[J]
[86]  C[2;]← 11 5 EFT VARSA[J]
[87]  C[3;]← 11 5 EFT S[J]
[88]  A[J;],'      ',C[1;],'      ',C[2;],'    ',C[3;]
[89]  →(J=7)/TAB8
[90]  →TAB7
[91] TAB8:→0
    ∇
```

```
      ∇EXPONP[⎕]∇
    ∇ EXPONP;X;Y;TT;SM;PC;BL;TL;PC2;R90;HS;GL;BT;PI;ST;D
[1]   'EXPONP ORDERS THE DATA YOU GIVE AND COMPUTES THE'
[2]   'EMPIRICAL  LOG  SURVIVER  FUNCTION FOR THE DATA.'
[3]   'A PLOT OF THE LOG SURVIVER FUNCTION FOR THE DATA'
[4]   'IS THEN PRINTED TO SEE IF THERE IS A LINEAR FIT.'
[5]   ' '
[6]   'IF YOU WANT TO TITLE  YOUR PLOT TYPE YOUR TITLE.'
[7]   'IF YOU DO NOT WANT A TITLE JUST HIT THE CARRIAGE'
[8]   'RETURN.'
[9]   ' '
[10]  TT←⍞
[11]  ' '
[12]  'ENTER YOUR DATA IN VECTOR FORM'
[13]  X←⎕
[14]  SM← 3 10
[15]  ' '
[16]  PC←'*+o'
[17]  BL←'ORDERED DATA'
[18]  TL←'EXPONENTIAL SCORES'
[19]  PC2←'*'
[20]  R90←HS←GL←0
[21]  D←130
[22]  ST← 1 1.25 1.5 2 2.5 3 4 5 7.5 10
[23]  PI←10
[24]  BT←10
[25]  X←X[⍋X]
[26]  Y←⍟((((⍴X)+1)-⍳(⍴X))÷((⍴X)+1))
[27]  (10 10 ,(⍴X)) MPLOT Y VS X
    ∇

      ∇ECDF[⎕]∇
    ∇ ECDF;XN;BN;FN;LIN;I;FMA;J;U;LOW;XINC;MAXM1;TRY
[1]   XN←(BN←((N)*0.5)÷C[12])÷N
[2]   LOW←I←1
[3]   FMA←¯10
[4]   XINC←HSCALE÷(A[1]×4)
[5]   FN←(MAXM1←((A[1]×4)-1))⍴0
[6]  TAB60:→(I>MAXM1)/TAB50
[7]   U←A[2]+I×XINC
[8]   J←LOW
[9]  TAB61:→(J>N)/TAB53
[10]  TRY←BN×(U-X[J])
[11]  →(TRY>1)/TAB52
[12]  →(TRY<¯1)/TAB53
[13]  FN[I]←FN[I]+1-|TRY
[14]  J←J+1
[15]  →TAB61
[16] TAB52:LOW←J
[17]  J←J+1
[18]  →TAB61
[19] TAB53:FN[I]←FN[I]×XN
[20]  FMA←FMA⌈FN[I]
[21]  I←I+1
[22]  →TAB60
[23] TAB50:I←1
[24] TAB62:→(I>MAXM1)/TAB54
[25]  LIN←⌊((D[1]+0.5)-(D[1]-1)×(FN[I]÷FMA))
[26]  ARRAY[LIN;I]←'P'
[27]  I←I+1
[28]  →TAB62
[29] TAB54:→0
    ∇

      ∇NDTRI[⎕]∇
    ∇ R←NDTRI P;W
[1]   R← 0 0
[2]   →((0.5≤|P-0.5),0=|P-0.5)/L1,L2
[3]   W←(⍟÷(P⌊1-P)*2)*0.5
[4]   L6:R[1]←(W-((+/(2.515517 0.802853 0.010328)×W*¯1+⍳3)÷+/(1 1.432788 0.189269 0.001308)×W*¯1+⍳4))××P-
      0.5
[5]   R[2]←0.3989423×*-0.5×R[1]×R[1]
[6]   →0
[7]   L1:R[1]←(×P-0.5)×10*74
[8]   →0
[9]   L2:R[2]←0.3989423
    ∇
```

```
    ∇NORMP[⎕]∇
  ∇ NORMP;X;TT;BL;SM;TL;PC2;I;J;R;R90;HS;GL;BT;PI;ST;D;PC
[1]   'NORMP ORDERS THE DATA YOU GIVE  AND  COMPUTES THE'
[2]   'INVERSE OF THE UNIT NORMAL  CUMULATIVE  DISTRIBU-'
[3]   'TION FOR THE DATA.  A PLOT OF THE INVERSE  OF THE'
[4]   'UNIT NORMAL CUMULATIVE DISTRIBUTION VS THE ORDER-'
[5]   'ED DATA IS THEN  PRINTED  TO  SEE  IF  THERE IS A'
[6]   'LINEAR FIT.'
[7]   ' '
[8]   'IF YOU  WANT TO TITLE YOUR PLOT TYPE YOUR  TITLE.'
[9]   'IF YOU DO NOT WANT A TITLE JUST HIT  THE CARRIAGE'
[10]  'RETURN.'
[11]  ' '
[12]  TT←⍞
[13]  ' '
[14]  'ENTER YOUR DATA IN VECTOR FORM'
[15]  X←⎕
[16]  ' '
[17]  BL←'ORDERED DATA'
[18]  SM← 3 10
[19]  R90←HS←GL←0
[20]  D←130
[21]  ST← 1 1.25 1.5 2 2.5 3 4 5 7.5 10
[22]  PI←10
[23]  BT←⍳0
[24]  TL←'NORMAL SCORES'
[25]  PC2←''
[26]  PC←'*-|'
[27]  X←X[⍋X]
[28]  R←((⍴X),2)⍴0
[29]  I←(⍳(⍴X))÷((⍴X)+1)
[30]  J←0
[31] TAB3:J←J+1
[32]  R[J;]←NDTRI I[J]
[33]  →(J=(⍴X))/TAB2
[34]  →TAB3
[35] TAB2:J←(,R[;1]),101⍴0
[36]  I←X,X[1],(X[1]+((X[(⍴X)]-X[1])÷100)×⍳100)
[37]  →(X[1]<0∧X[(⍴X)]>0)/TAB4
[38]  (10 10 ,(⍴X),(101)) MPLOT J VS I
[39]  →TAB5
[40] TAB4:J←J,R[1;1],(R[1;1]+((R[(⍴X);1]-R[1;1])÷48)×⍳48)
[41]  I←I,49⍴0
[42]  (10 10 ,(⍴X),(101),(49)) MPLOT J VS I
[43] TAB5:→0
  ∇
```

```
    ∇AUTOS[⎕]∇
  ∇ AUTOS
[1]   →((⍴X)≥80)/TAB70
[2]   A[1]←16
[3]   →TAB71
[4]  TAB70:A[1]←⌊(28⌊(⍴X÷5))
[5]  TAB71:A[2]←C[21]
[6]   A[3]←C[27]
[7]   →0
  ∇
```

```
    ∇EFT[⎕]∇
  ∇ Z←W EFT X;D;E;H;J;K;L;Q;S;T;U;Y
[1]   D←'0123456789.E ¯'
[2]   →(∨/W≠⌊W←,W+(H←0)×L←1<⍴⍴X)/EFTERR+0×K←2
[3]   →(3 2 1 <⍴⍴X)/(EFTERR+K←0), 2 3 +I26
[4]   X←((∨/ 1 2 =⍴W)⌽ 1 2)⍉(1,⍴,X)⍴X
[5]   X←(⌽2⍴⌽⍴X)⍴X
[6]   →((∧/(⍴W)≠ 1 2 ,2×E←1⍴⌽⍴X),1≠⍴W)/(EFTERR×K←1),2+I26
[7]   W←(W+6+(∨/,X<0)+∨/,1>|X),W
[8]   →(∨/6>-/[1] W←⍉(E,2)⍴W)/EFTERR+0×K←2
[9]   Z←((K←1⍴⍴X),+/W[1;])⍴' '
[10] EFTLP:→(E<H←H+1)/EFTEND
[11]  S←1+⌊10⍟|Y+0=Y←X[;H]
[12]  U←1+⌊10⍟|Y+0=Y←⌊0.5+(10*Q-15)+Y×10*(Q←W[2;H])-S
[13]  J←(((T-4)⍴1),4⍴0)\1+⌊10|(|Y÷10*U>Q)∘.÷10*¯1+⌽⍳¯4+T←W[1;H]
[14]  J[;T- 2 1]←1+⌊10|(|S-U≤Q)∘.÷ 10 1
[15]  J[;(⍳U+T-4+Q),T]←13
[16]  J[;1,U,T,T-3]←⍉(4,K)⍴(K⍴11),(13+0>Y,S-1),K⍴12
[17]  J[;⍳T-3]←J[;(1⌽⍳U+1),(U+1+⍳Q)]
[18]  J[;T- 2 1 0]←(-S≤0)⌽J[;T- 2 1 0]
[19]  →EFTLP,⍴Z[;(+/W[1;⍳H-1])+⍳T]←D[J]
[20] EFTEND:→L/0
[21]  →0×⍴Z←,Z
[22] EFTERR:'EFT ',(3 6 ⍴' RANK LENGTHDOMAIN')[K+1;],' PROBLEM.'
  ∇
```

```
      ∇APLOT[⎕]∇
    ∇ APLOT;I;J;LINE;CROB;PROB;VERT;H1;PLABEL;DIB;FSCALE;DID;DIT;DIS;IQ1T;IQ2T;IQ3T;NMAX;MNT;RT;PRBMX;INCR;ARRAY;FMAX
[1]    →((ρTT)=(ρι0))/TAB5A
[2]    OL OUT(18ρ' '),TT
[3]    OL OUT 1ρ' '
[4]   TAB5A:ARRAY←((D[1]),(4×A[1]))ρ' '
[5]    FSCALE←(D[1]-1)÷FMAX←(F[⍋F])[ρF]
[6]    I←0
[7]    J←¯3
[8]   TAB12:→(I=A[1])/TAB15
[9]    I←I+1
[10]   J←J+4
[11]   LINE←⌊((D[1]-0.5)-(FSCALE×F[I]))
[12]   →(LINE≥(D[1]-1))/TAB13
[13]   ARRAY[LINE+ι(D[1]-LINE);J]←'*'
[14]   ARRAY[LINE+ι(D[1]-LINE);J+1]←'*'
[15]   ARRAY[LINE+ι(D[1]-LINE);J+2]←'*'
[16]   ARRAY[LINE+ι(D[1]-LINE);J+3]←' '
[17]   ARRAY[(D[1]);J+3]←'+'
[18]   →TAB12
[19]  TAB13:→(F[I]≠0)/TAB14
[20]   ARRAY[(D[1]);J]←'-'
[21]   ARRAY[(D[1]);J+1]←'-'
[22]   ARRAY[(D[1]);J+2]←'-'
[23]   ARRAY[(D[1]);J+3]←'+'
[24]   →TAB12
[25]  TAB14:ARRAY[(D[1]);J]←'*'
[26]   ARRAY[(D[1]);J+1]←'*'
[27]   ARRAY[(D[1]);J+2]←'*'
[28]   ARRAY[(D[1]);J+3]←'+'
[29]   →TAB12
[30]  TAB15:PROB←((D[1]),4)ρ' '
[31]   INCR←(PRBMX←FMAX÷N)÷9
[32]   CROB←PRBMX,(PRBMX-INCR×ι8),0
[33]   CROB← 4 2 DFT CROB← 10 1 ρCROB
[34]   PROB[D[2]+D[3]×ι10;]←CROB[ι10;]
[35]   VERT←((D[1]),1)ρ'|'
[36]   RT←(NMAX←A[1]×4)÷(A[3]-A[2])
[37]   IQ1T←⌊(0.5+(C[23]-A[2])×RT)
[38]   IQ2T←⌊(0.5+(C[24]-A[2])×RT)
[39]   IQ3T←⌊(0.5+(C[25]-A[2])×RT)
[40]   MNT←⌊(0.5+(C[1]-A[2])×RT)
[41]   →(MNT≥1)/TAB21
[42]   MNT←1
[43]  TAB21:→(MNT≤NMAX)/TAB22
[44]   MNT←NMAX
[45]  TAB22:→(IQ1T≥1)/TAB23
```

```
 IQ1T←1
TAB23:→(IQ2T≥1)/TAB24
 IQ2T←1
TAB24:→(IQ3T≥1)/TAB25
 IQ3T←1
TAB25:→(IQ1T≤NMAX)/TAB26
 IQ1T←NMAX
TAB26:→(IQ2T≤NMAX)/TAB27
 IQ2T←NMAX
TAB27:→(IQ3T≤NMAX)/TAB28
 IQ3T←NMAX
TAB28:ARRAY[⍳(D[1]);IQ1T]←'.'
 ARRAY[⍳(D[1]);IQ2T]←'.'
 ARRAY[⍳(D[1]);IQ3T]←'.'
 ARRAY[⍳(D[1]);MNT]←'M'
 ARRAY[⍳(D[1]);NMAX]←'|'
 →(B=0)/TAB3A
 ECDF
TAB3A:H1←(5⍴' '),H1←'FREQUENCIES',H1←(32⍴' '),H1←'SAMPLE SIZE = '
 OL OUT H1, 6 0 DFT(⍴X)
 OL OUT 1⍴' '
 OL OUT H1←(4⍴' '),H1← 4 0 DFT F
 OL OUT '     ','+',(4×A[1])⍴'---+'
 OL OUT '     ','|',H1←((4×(A[1]-1))⍴' '),H1←'   |'
 OL OUT PROB,VERT,ARRAY
 DIS←⌊((⍴XLABEL)÷2)
 DIT←(⍴XLABEL)÷2
 →((DIT-DIS)=0)/TAB40
TAB41:DID←(8×DIS)⍴'|       '
 OL OUT '     ',DID,'|'
 →TAB42
TAB40:DIS←DIS-1
 →TAB41
TAB42:DIB←DIS+1
 XLABEL←XLABEL[¯1+2×⍳DIB]
 →(⌈XLABEL>99999)/TAB31
 →(⌊XLABEL<¯9999)/TAB31
 →(DELTA<0.1)/TAB31
 PLABEL←((⍴XLABEL),1)⍴XLABEL
 OL OUT ' ',PLABEL←,PLABEL←(PLABEL← 7 1 DFT PLABEL),((⍴XLABEL),1)⍴' '
 →0
TAB31:XLABEL←XLABEL[¯1+2×⍳(⌈((⍴XLABEL)÷2))]
 PLABEL←((⍴XLABEL),1)⍴XLABEL
 OL OUT ' ',PLABEL←,PLABEL←(10 4 EFT PLABEL),((⍴XLABEL),6)⍴' '
 →0
∇
```

```
      ∇MULTIPLOT[⎕]∇
    ∇ MULTIPLOT I;J;L;T;PT;U;K;M;N;L̲;L1;L2;L3;L4;L5;E;TM;HM;Q;Q1;Q8;R
[1]    D̲←D̲⌊1↑C,6⍴3,C←6⍴ 3 120
[2]    MSGS 'OFF'
[3]    →(~R90)/PL2,ST←6⍴K←R←0
[4]    (S̲M̲[2]-⍴,K←H[1]) TICMARK⍴⍴P[2]←'¯'
[5]   PL2:L̲←(1,G̲L̲∧.∧HM←0=(S̲M̲[2]↑ 1 2)∘.|⍳H[2])\2
[6]    8 TICMARK⍴L3←PL3↑1↑HS,L2←P̲×⍳Q←~HS,⍴C←H[1]
[7]    L5←P̲-1↑HS,L4←Q̲×⍳~P2,⍴M←(⍴P)[1]-1
[8]    L1←((HS,HS∨A̲←0≥Q8←D̲-8)∧~A̲)/PL4↑ ¯2 1
[9]    TM←TM,[1.5] TM↑ ¯1↑TM
[10]  PL3:E←~I←0∧×L←L̲×R
[11]   →(L1×⍳Q<⍴L[D←1↑D/X[;2]]←J←2↑(D←X[;1]=N←|K-C)/A),L2
[12]   →P̲,L[T]←(E\L̲)[T←(R∧L≤1)/⍳⍴L]
[13]   D←(E/⍳⍴E)[T,(U←L̲[T]>2)/T←D]
[14]   L̲[(~U)/T]←(~U,U/1)/J←J,U/M+1
[15]  PL4:→(A̲∨1≥⍴D)/P̲,E←1∨×L←L̲
[16]   →(L5×⍳~∨/U←(P2∨T←J≠1⌽J←1,J[T])∧D=1⌽D←0,D[T←U[⍋D[U←⍋J]]]),L4
[17]   →(∧/U←1=+/(D∘.=D←U/D)∧J∘.=J)/Q̲,J[T/⍳⍴J←U/J]←(T←U/~T)/M-1
[18]   U←1↓(T=1⌽T←1,J←J[T])∧U=1⌽U←0,D←D[T←U[⍋D[U←⍋J]]]
[19]  Q̲:→P̲×⍳Q8<⍴T←~(⍳T↑T←⍴I)∊D←(D-1),D←D+2×⍳T←⍴D←U/D
[20]   I←T\I
[21]   I[D]←(U/M),U/J
[22]   L←I⌈(E←~×I)\L
[23]  P̲:→(×N)/PL5←∨/T←(2 1 ×N)∊,TM
[24]   →(PL5+1),L←L⌈E\1↑1,∨⌿HM
[25]  PL5:L←L⌈0,T,((¯4+⍴L)⍴G̲L̲∧1↑T),0
[26]   PT[TM[;1]⍳N;],P[1+L]
[27]   →(0≤R←×C←C-1)/L3
[28]   (S̲M̲[2]-1) TICMARK-R90
[29]   →U←(ST[3 4],1)/ 1 3 4 +⍴26
[30]   'SCALE FACTOR FOR ORDINATE: ';10*ST[5]
[31]   →U←1↓U
[32]   'SCALE FACTOR FOR ABSCISSA: ';10*ST[6]
[33]   MSGS 'ON'
    ∇
```

```
      ∇DFT[⎕]∇
    ∇ Z←W DFT X;D;E;F;G;H;I;J;K;L;Y
[1]   D←' 0123456789.-'
[2]   →(∨/W≠⌊W←,W+(H←0)×L←1<⍴⍴X)/DFTERR+0×F←2
[3]   →(3 2 1 <⍴⍴X)/(DFTERR+F←0), 2 3 +⌶26
[4]   →(2+⌶26),⍴X←((∨/ 1 2 =⍴W)⌽ 1 2)⍉(1,⍴,X)⍴X
[5]   X←(0 1 1 /⍴X)⍴X
[6]   →((∧/(⍴W)≠ 1 2 ,2×E←1⍴⌽⍴X),1≠⍴W)/(DFTERR×F←1),3+⌶26
[7]   I←1+⌈/0,,⌊10⍟|X+1>|X
[8]   W←(2+I+W+(W≠0)+∨/,X<0),W
[9]   →(∨/2>-/[1] W←⍉(E,2)⍴W)/DFTERR+0×F←2
[10]  Z←((K←1⍴⍴X),+/W[1;])⍴' '
[11]  X←⌊0.5+X×10*(⍴X)⍴W[2;]
[12] DFTLP:→(E<H←H+1)/DFTEND
[13]  J←1+⌊10|(|Y←X[;H])∘.÷10*¯1+⌽⍳I←W[1;H]
[14]  J←(,J)×G←,⍉(⌽⍴J)⍴(,⍉(J≠1)∨.∧(⍳I)∘.≤⍳I-F+1),(K×1+F←W[2;H])⍴1
[15]  →(∧/0≤Y)/2+⌶26
[16]  J[1+(⍴J)|¯1+(I-+/(K,I)⍴G)+I×¯1+⍳K]←12×Y<0
[17]  J←(K,I)⍴J
[18]  →(0=F)/3+⌶26
[19]  J←J[;(1⌽⍳G),(G←-/W[;H])+⍳F]
[20]  J[;G]←11
[21]  →DFTLP,⍴Z[;(+/W[1;⍳H-1])+⍳I]←D[1+J]
[22] DFTEND:→L/0
[23]  →0×⍴Z←,Z
[24] DFTERR:'DFT ',(3 6 ⍴' RANK LENGTHDOMAIN')[F+1;],' PROBLEM.'
    ∇


      ∇AUTOSCALE[⎕]∇
    ∇ AUTOSCALE;C;D
[1]   C←C+(X[1;]+X[1;]=0)×0=C←(⌈/X)-D←⌊/X
[2]   F←F÷G←10*⌊10⍟F←|C÷H←SM×⌊((6,PI)×2⍴A)÷SM←⌈16⌊SM⌈ 1 5
[3]   F←G×ST[+/F∘.≤ST←10,ST[⍒ST]]
[4]   X←(C⍴ 0 0.5)+(C⍴÷F)×X-(C+⍴X)⍴G←G×(0<G-C+C)∨0>G←C×⌊D÷C←F×SM÷2
    ∇
```

```
      ∇APLNAME[⎕]∇
    ∇ FID←APLNAME A;K;REM
[1]   ⍝ REMOVE EXTRA BLANKS
[2]   A←1↓(K∨1⌽K←' '≠A)/A←' ',A,' '
[3]   ⍝ FIND END OF FILENAME
[4]   K←(A⍳' ')-⍳1
[5]   ⍝ IF ONE WORD - SYNTAX ERROR
[6]   →(K=⍴A)/ER1
[7]   ⍝ EXTRACT FILENAME
[8]   FID←8↑K↑A
[9]   ⍝ AND REMAINDER
[10]  REM←(K+1)↓A
[11]  ⍝ FIND END OF FILETYPE
[12]  K←(REM⍳' ')-⍳1
[13]  ⍝ ADD FILETYPE TO FILE
[14]  FID←FID,(8↑K↑REM)
[15]  ⍝ EXTRACT 2ND REMAINDER
[16]  REM←(K+1)↓REM
[17]  ⍝ CHECK SPECIAL MODES
[18]  →((∧/'SY'=2↑REM)∨(∧/'* '=2↑REM))/L1
[19]  ⍝ MODELETTER='P' UNLESS OTHERWISE
[20]  FID←FID,'ABCTP'['ABCT'⍳1↑REM]
[21]  ⍝ MODENO='1' UNLESS OTHERWISE
[22]  FID←FID,'0234561'['023456'⍳1↑1↓REM]
[23]  →L2
[24] L1:FID←FID,2↑REM
[25]  ⍝ RECTYPE='F' UNLESS V SPECIFIED
[26] L2:FID←FID,' ','FV'[('V'=¯1↑REM)+⍳1]
[27]  ⍝ CONVERT TO EBCDIC INTEGER
[28]  FID←2 OF FID
[29]  →0
[30] ER1:'FILETYPE MISSING'
    ∇
```

```
     ∇MPLOT[⎕]∇
   ∇ A MPLOT X;C;D;F;G;H;P;P2;HS;Δ;ST
[1]   INITIAL
[2]   AUTOSCALE
[3]   SETΔAP
[4]   MULTIPLOT H←S̲M̲×⌈((⌈/X←⌊X)÷S̲M̲
   ∇


     ∇OUT[⎕]∇
   ∇ OL OUT R;I;J;MAX
[1]   →(2=ρρR)/L1
[2]   R←(1,ρR)ρR
[3]  L1:→(OL=1)/OFF
[4]   →0,ρ⎕←R
[5]  OFF:MAX←1↑ρR←' ',R
[6]   J←20⌈¯1↑ρR
[7]   I←1
[8]  L2:(J↑R[I;]) WRITE APLN
[9]   →(MAX≥I←I+1)/L2
   ∇


     ∇INITIAL[⎕]∇
   ∇ INITIAL
[1]   →((0=×/(2ρA),D←ρX), 2 1 <ρρX)/0,PL2- 1 0
[2]   →PL2,D←ρX←(ιρ,X),[1.5] X
[3]   X←(D←¯2↑D)ρX
[4]  PL2:X←R90⌽(,⍉ 0 1 ↓X),[1.5](C←×/D←D- 0 1)ρX[;1]
   ∇
```

```
     ∇AND[⎕]∇
   ∇ L←A AND B;C;D
[1]   →(((2<ρρA)∨3<ρρB),0≠ρρB)/ 17 3
[2]   B←,B
[3]   →(((3=ρρB)∧1≠1ρρB),2=ρρA)/ 17 7
[4]   A←,A
[5]   →(∧/((ρA)≠1,D),1≠D←1ρ¯2⌽ρB)/16
[6]   A←(((D×ρA)⌊D⌈ρA),1)ρA
[7]   →(1≠ρρB)/9
[8]   B←(((ρB)⌈(1=ρB)×1ρρA),1)ρB
[9]   →((∧/D≠1,1ρρA),1≠D←1ρ¯2⌽ρB)/ 16 11
[10]  B←(((3=ρρB)ρ1),(1ρρA),1ρ⌽ρB)ρB
[11]  →(3=ρρB)/14
[12]  L←(((C←1ρ⌽ρA)ρ0),(1ρ⌽ρB)ρ1)\B
[13]  →0×ρρL[;ιC]←A
[14]  L←(1,((C←1ρ⌽ρA)ρ0),(¯1+1ρ⌽ρB)ρ1)\B
[15]  →0×ρρL[;;1+ιC]←A
[16]  →0=ρ⎕←'ARGUMENTS OF  AND  ARE NOT CONFORMABLE.'
[17]  'AN ARGUMENT OF  AND  IS OF IMPROPER RANK.'
   ∇


     ∇TOT[⎕]∇
   ∇ TOT
[1]   ARRAY[1;]←,BRRAY[1+ι(A-1);]
[2]   K←1
[3]  TAB1:K←K+1
[4]   ARRAY[K;]←((,BRRAY[K+ι(A-K);])),(,BRRAY[ι(K-1);])
[5]   →(K=(A-1))/TAB2
[6]   →TAB1
[7]  TAB2:ARRAY[A;]←,BRRAY[ι(A-1);]
[8]   →0
   ∇
```

```
      ∇TICMARK[⎕]∇
    ∇ U TICMARK ISV;C;I;J;L;T;NB;VT;O;N;E;K
[1]   →(PL3-⍳~K←R90≠ISV),BP×⍳~ISV∊ 0 1
[2]   (×⍴,TT)/BS[2],((0⌈⌊0.5×8+H[2]-⍴,TT)⍴' '),TT,1↓BS
[3]   ((R90×0⌈8+⌊0.5×H[2]-⍴,TL)⍴' '),TL
[4]   C←⍴NB←G[J]+F[J]×TM←0,SM[J]×⍳⌊H[J]÷SM[J+2-ISV]
[5]  PL3:N←(C,VT←∨/N)⍴(N←0>PT←⌈NB×10*U-I←⌈/O←1⌈1+E←⌊10⍟|NB+NB=0)\'-'
[6]   L←(U←U-N←-VT)+1-VT+(⌽'0'∧.=PT←N,'0123456789'[1+⍉(U⍴10)⊤|PT])⍳0
[7]   →((U>T←VT+⌈/I,(L←L≠I),(I≤0)×2+L-I),L≥U-VT+L>I)/ 3 2 +⌶26
[8]   →(⌶26)-I←⍴,ST[6-K]←I-1
[9]   →PL3,⍴NB←(10*-L)×⌊0.5+NB×10*(L←¯3+U+U-VT)-ST[6-K]←⌈/(NB≠0)/E
[10]  PT←(~(⍳U+J)∊(I+J+VT),⍳¯1+J+U-T)\(C,U)⍴(,VT⌽O)\(,O←(I-O)∘.<N⌽⍳U)/,PT
[11]  ST[4-K]←ST[6-K]≠⍴⍴PT[;I+J+VT]←'.'
[12]  →ISV/0×⍴PT←(((C+1),~VT)⍴' '),(1 ¯1 +C,U)↑PT
[13]  (SM[2]-9)⌽,' ',PT
[14]  →R90/0
[15] BP:(((~R90)×0⌈8+⌊0.5×H[2]-⍴,BL)⍴' '),BL
[16]  (×⍴,BT)/BS[2],((0⌈⌊0.5×8+H[2]-⍴,BT)⍴' '),BT,1↓BS
    ∇


      ∇SETΔAP[⎕]∇
    ∇ SETΔAP
[1]   D←⍴A←(A>A←0)/A←A,C←+/A←2↓A,D[2]⍴D[1]
[2]   →(D>1)/4,⍴P←' _|',(D⍴PC),((P2←(×⍴,PC2)∧~HS←HS∧~R90)↑PC2),1↓BS
[3]   →~A←⍴A←~P2
[4]   A←1++/(⍳C)∘.>(D∘.≥D←⍳D)+.×A
    ∇


      ∇VS[⎕]∇
    ∇ M←A VS B;C;D
[1]   →(((⍴⍴B←,B)<⍴⍴B), 2 1 0 <⍴⍴A)/ 8 8 4 3
[2]   A←((⍴B),1)⍴A
[3]   A←((×/⍴A),1)⍴A
[4]   →(∧/(⍴B)≠1,1⍴⍴A)/9
[5]   M←(0,(1⍴⌽⍴A)⍴1)\A
[6]   M[;⍳1]←B
[7]   →0×⍴⍴M←(1,⍴M)⍴M
[8]   →0=⍴⎕←'AN ARGUMENT OF  VS  IS OF IMPROPER RANK.'
[9]   'ARGUMENTS OF  VS  ARE NOT CONFORMABLE.'
    ∇
```

AD-A035 883 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 9/2
APL HISTOGRAM, DENSITY ESTIMATION AND PROBABILITY PLOTTING ROUT--ETC(U)
DEC 76 D R HUTCHINSON

UNCLASSIFIED NL

2 OF 2
ADA035883

END
DATE
FILMED
3 - 77

LIST OF REFERENCES

1. Chatfield, C. and P. Sprent. "Convenient Probability Plotting Positions for the Normal Distribution," Applied Statistics 25: 47-50, 1976.

2. Dixon, M. W. A Statistical Analysis of Deviation from Target Cost in NAVAIRSCOMHQ Fixed-Price Incentive Contracts during the 1949-1965 Time Frame. U.S. Naval Postgraduate School, Monterey, California, 1973.

3. Fishman, G. S. Concepts and Methods of Discrete Event Digital Simulation. John Wiley and Sons, Inc., New York, 1973.

4. Gillman, L. and A. J. Rose. APL An Interactive Approach. John Wiley and Sons, Inc., New York, 1974.

5. Lewis, P.A.W. "A Branching Poisson Process Model for the Analysis of Computer Failure Patterns." The Journal of the Royal Statistical Society Series B (Methodological). 26: 398-456, 1964.

6. Lewis, P.A.W. and D. R. Cox. "A Statistical Analysis of Telephone Circuit Error Data." IEEE Transactions on Communications Systems, Com-14, 382-389, 1966.

7. Lewis, P.A.W. and G. S. Shedler. "Empirically Derived Micromodels for Sequences of Page Exception." IBM Journal of Research and Development 17: 89-90, 1973.

8. Lewis, P.A.W., L. H. Liu, D. W. Robinson and M. Rosenblatt. Empirical Sampling Study of a Goodness of Fit Statistic for Density Function Estimation. U. S. Naval Postgraduate School, Monterey, California, 1975.

9. Miller, R. G. "The Jacknife - A Review." Biometrika 61: 1, 1974.

10. Pearson, E. S. and H. W. Hartley. Biometrika Tables for Staticians, v. I. Cambridge University Press, Cambridge, 1966.

11. Rosenblatt, M. "Curve Estimates." The Annals of Mathematical Statistics, v. 27, 1956.

12. Weast, R. C. C.R.C. Standard Mathematical Tables, 19th Ed., The Chemical Rubber Co., Ohio, 1971.

13. Wilk, A. and B. Gnanadesikan. "Probability Plotting Methods for the Analysis of Data." Biometrika 55: 1-17, 1968.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Chief, Civil Schools Branch Department of Army U. S. Army Military Personnel Center Alexandria, Virginia 22332	1
3.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
4.	Dr. P.A.W. Lewis, Code 55Lw Department of Operations Research Naval Postgraduate School Monterey, California 93940	15
5.	Lt. D. R. Bouchoux, Code 55Bh Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
6.	Dr. R. W. Butterworth, Code 55Bd Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
7.	Dr. D. T. Gavor, Code 55Gv Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
8.	Dr. K. T. Marshall, Code 55Mt Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
9.	Dr. R. R. Read, Code 55Re Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
10.	Capt. Dennis R. Hutchinson 1333 Capistrano Drive Salt Lake City, Utah 84116	2